THE ADVANCED THEORY OF STATISTICS

by

MAURICE G. KENDALL, M.A.

Fellow and Member of the Council of the Royal Statistical Society; Statistician to the Chamber of Shipping of the United Kingdom; formerly Head of the Economics Intelligence Branch, Ministry of Agriculture and Fisheries

VOLUME

With 16 Illustrations and 79 Tables

LONDON

CHARLES GRIFFIN & COMPANY LIMITED

42 DRURY LANE

1943

Printed in Great Britain
by Butler & Tanner Limited, Frome

DEDICATED

TO MY MOTHER.

"Let us sit on this log at the roadside," says I, "and forget the inhumanity and ribaldry of the poets. It is in the glorious columns of ascertained facts and legalised measures that beauty is to be found. In this very log we sit upon, Mrs. Sampson," says I, "is statistics more wonderful than any poem. The rings show it was sixty years old. At the depth of two thousand feet it would become coal in three thousand years. The deepest coal mine in the world is at Killingworth, near Newcastle. A box four feet long, three feet wide, and two feet eight inches deep will hold one ton of coal. If an artery is cut, compress it above the wound. A man's leg contains thirty bones. The Tower of London was burned in 1841."

"Go on, Mr. Pratt," says Mrs. Sampson. "Them ideas is so original and soothing. I think statistics are just as lovely as they can be."

O. Henry, *The Handbook of Hymen.*

widespread applications of χ^2 in testing agreement between theory and observation I felt that it should be introduced at an early stage.

The second volume will deal with the Theory of Estimation, Regression, Analysis of Variance, Tests of Significance, Multivariate Analysis, Theories of Statistical Inference, and Time Series. In the first volume it has been possible to avoid a detailed examination of controversial topics connected with the logic of inference in probability ; the subject will be taken up more systematically in the second volume.

On the invaluable principle that example is better than precept, a special effort has been made to exemplify the theory at every stage and to provide exercises for the reader to work out for himself. Some of the latter are rather difficult, but have nevertheless been included to illustrate the scope of application of the theory and to refer to results for which no place could conveniently be found in the text. In assembling this material I have drawn freely on the wealth of research work in statistical periodicals, particularly *Biometrika*, and am glad to make acknowledgment to the authors from whose papers examples have been taken.

Foremost among my more specific indebtedness is that to Dr. Leon Isserlis, who read the whole book at the galley proof stage and to whose careful scrutiny I owe a great deal. I have also to thank Dr. J. O. Irwin, who allowed me to consult his draft of a chapter originally intended for the co-operative treatise (this forms the basis of Chapter 10); Professor R. A. Fisher and Messrs. Oliver and Boyd, for permission to reproduce Appendix Tables 4 and 5 from the former's *Statistical Methods for Research Workers* ; and the publishers, Messrs. Charles Griffin and Co., and the printers, Messrs. Butler & Tanner Ltd., who have taken great pains with some very difficult manuscript.

I shall be grateful to any reader who notifies me of any error, omission or ambiguity, from which, I fear, no book of this kind can be entirely free at its first appearance.

M. G. KENDALL.

LONDON,
February 1st, 1943.

TABLE OF CONTENTS

INTRODUCTORY NOTE

0.1. The chapter-sections in this book are numbered serially. The serial numbers are prefixed by the number of the chapter in which they occur and are separated therefrom by a period, e.g. **14.13** refers to the thirteenth section of Chapter 14. A similar procedure is followed for tables, equations and exercises, e.g. (7.15) refers to the fifteenth equation of Chapter 7. In cross-references, chapter-sections are denoted by clarendon type, the others by ordinary type.

0.2. References to printed work are given by author's name and date of publication. In the list of references at the end of the chapter authors are arranged alphabetically. Where articles from publications are referred to, the number of the volume is given in clarendon type and the number of the first page of the article in ordinary type, e.g. *Ann. Math. Statist.*, **10**, 275, refers to the article beginning on page 275 of volume 10 of the *Annals of Mathematical Statistics*. Where an exercise is followed by an author's name and a date, the result given in the exercise appears in the article listed in the references to the chapter concerned under these particulars. Where the result is from an article not previously referred to a full reference is given.

0.3. The mathematical notation is that in current use, but a few symbols may be explained.

(1) The exclamation mark ! written after an integer means the factorial of that integer. Some writers give the symbol a more extended use for non-integral numbers by writing

$$x! = \Gamma(x+1) = \int_0^\infty e^{-t} t^x \, dt.$$

This, of course, accords with the factorial notation, but will not be used in this book.

(2) The combinatorial sign $\binom{n}{r} = \dfrac{n!}{r!(n-r)!}$ will be used in place of the older ${}^nC_r$.

(3) The summation sign will be written as Σ, e.g. $\sum_{j=1}^{j=n} x_j = x_1 + x_2 + \ldots + x_n$.

The symbol $\sum_{j=1}^{j=n}$ can as a rule be shortened to $\sum_{j=1}^{n}$ and in many cases to $\sum_j$ or merely to Σ, the extent of the summation being clear from the context.

(4) The ordinary notation for the Γ-function (given above), the B-function, and the hypergeometric function will be used, i.e.

$$B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}$$

and

$$F(\alpha,\beta,\gamma,x) = 1 + \frac{\alpha.\beta}{1.\gamma}x + \frac{\alpha(\alpha+1).\beta(\beta+1)}{1.2.\gamma(\gamma+1)}x^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1.2.3.\gamma(\gamma+1)(\gamma+2)}x^3 + \ldots$$

(5) Where the exponent is concise, the exponential function will be written as a power of e, for example $e^{\frac{1}{2}x^2}$. But where it is lengthy we shall use the notation exemplified by $\exp\{-\frac{1}{2}(x^2 - 2\rho xy + y^2)\}$ instead of $e^{-\frac{1}{2}(x^2-2\rho xy+y^2)}$.

0.4. In some fields it is useful to preserve a distinction between a statistical parameter in a population and the estimate of that parameter from a sample. Where possible, the former will be denoted by a Greek letter and the latter by a Roman letter, e.g. the product-moment correlation coefficient of a population is denoted by ρ and that of a sample by r. It is not, however, always possible to preserve this distinction, as for instance with the multiple correlation coefficient R, in which case a Greek capital would be confused with the Roman P. Complete notational consistence can only be achieved at the expense of jettisoning a great deal of accepted statistical usage, and even then would probably result in some cumbrous symbols.

0.5. In order to enable the reader to follow the worked examples and illustrative material, a few tables of functions commonly required are given at the end of this volume. These tables are in no way a substitute for the comprehensive sets which have been published and which are a necessary adjunct to most practical and a good deal of theoretical work. Frequent reference will be made to the following :—

Tables for Statisticians and Biometricians, edited by Karl Pearson, Parts I and II, Biometrika Office, University College, London, W.C.1.

Statistical Tables for use in Biological, Agricultural and Medical Research, by R. A. Fisher and F. Yates, Oliver and Boyd, Edinburgh.

The following are also useful :—

Tables of the Incomplete Γ-function, edited by Karl Pearson, Biometrika Office, University College, London, W.C.1.

Tables of the Incomplete B-function, edited by Karl Pearson, Biometrika Office, University College, London, W.C.1.

The Kelley Statistical Tables, by T. L. Kelley, Macmillan, London and New York.

"Tables of Pearson's Type III Function," by L. R. Salvosa, *Ann. Math. Statist.*, 1930, **1**, 191.

Tables of the Higher Mathematical Functions, edited by H. T. Davis, Parts I and II, Principia Press, Bloomington, Indiana.

Tables of Random Sampling Numbers, by L. H. C. Tippett, Tracts for Computers, No. 15, Cambridge University Press.

Tables of Random Sampling Numbers, by M. G. Kendall and B. Babington Smith, Tracts for Computers, No. 24, Cambridge University Press.

Tables of the Correlation Coefficient, by F. N. David, Biometrika Office, University College, London, W.C.1.

Tables of $\tan^{-1} x$ *and* $\log(1 + x^2)$, by L. J. Comrie, Tracts for Computers, No. 23.

Tables of the Probability Integral, by W. F. Sheppard, British Association Mathematical Tables, Vol. 7, Cambridge University Press.

0.6. The references given at the end of the chapters are mainly intended to guide further reading and are not exhaustive. A more complete bibliography will be found in Mr. Yule's and my *Introduction to the Theory of Statistics*, which contains about 700 references to work appearing up to about 1932, and in the valuable periodic reviews of recent advances in theoretical statistics appearing in the *Journal of the Royal Statistical Society* and the *Journal of the American Statistical Association*. A recently-begun monthly publication by the American Mathematical Society, *Mathematical Reviews*, also contains material of interest in this connection.

CHAPTER 1

FREQUENCY-DISTRIBUTIONS

Statistics as the Science of Populations

1.1. Among the many subjects about which statisticians disagree is the definition of their science. In the *Revue de l'Institut International de Statistique* for 1935 (vol. 3, page 388) Dr. W. F. Willcox listed well over a hundred definitions of statistics, and the list was far from exhaustive. Even when we exclude those definitions which were formulated before the subject reached its present extent we are left with a variety of choices, and there is no definitive description of the scope of the science of statistics with which we can begin this book.

1.2. The fundamental notion in statistical theory is that of the group or aggregate, a concept for which statisticians use a special word—" population ". This term will be generally employed to denote any collection of objects under consideration, whether animate or inanimate; for example, we shall consider populations of men, of plants, of mistakes in reading a scale, of barometric heights on different days, and even populations of ideas, such as that of the possible ways in which a hand of cards might be dealt. The notion common to all these things is that of aggregation.

It is with the properties of populations that statistics is mainly concerned. In considering a population of men we are not interested, statistically speaking, in whether some particular individual has brown eyes or is a forger, but rather in how many of the individuals have brown eyes or are forgers, and whether the possession of brown eyes goes with a propensity to forgery in the population. We are, so to speak, concerned with the properties of the population itself. Such a standpoint can occur in physics as well as in demographic sciences. For instance, in discussing the behaviour of a gas we are not so much interested in the behaviour of particular molecules, as in that of the aggregate of molecules which go to compose the gas. The statistician, like Nature, is mainly concerned with the species and is careless of the individual.

1.3. We may therefore begin an approach to a definition of our subject by the following: statistics is the branch of scientific method which deals with the properties of populations. This, however, is rather too general. Statistics deals only with the *numerical* properties. A dictionary, for example, sets out a population of words, and among the properties of that population which are a suitable subject for scientific inquiry is that of word-derivation. It is not of statistical concern, however, to know that some words are derived from Latin, some from Anglo-Saxon and some from Hindustani. The subject would only assume a statistical aspect if we were to inquire *how many* words were derived from the different sources.

1.4. As a second approximation to our definition we may then try the following: statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of populations.

This again is a little too general. A set of logarithm tables is a population of numerals, but it is hardly a subject for statistical inquiry, for every numeral is determined according

to mathematical laws. The statistician is rather concerned with populations which occur in Nature and are thus subject to the multitudinous influences at work in the world at large. His populations rarely, if ever, conform exactly to simple mathematical rules, and in fact it is in the departure from such rules that he often finds topics of the greatest statistical interest. To allow for this factor we may then formulate our definition as follows :—

Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of populations of natural phenomena. In this definition " natural phenomena " includes all the happenings of the external world, whether human or not.

This is as far as we need pursue the matter. The reader who is interested enough to look through the definitions listed by Dr. Willcox in the article referred to above will find, I think, that in the light of this definition there is a perceptible thread of continuity running through them.

1.5. For the avoidance of misunderstandings in the interpretation of this definition it may be as well to point out that " statistics," the name of the scientific method, is a collective noun and takes the singular. The same word " statistics " is also applied to the numerical material with which the method operates, and in such a case takes the plural. Later in this book we shall meet the singular form " statistic," which is not, as might be supposed, an individual item of information which in the aggregate would compose " statistics," but is the name given to an estimate of certain unknown measures of a population.

Frequency-Distributions

1.6. Consider a population of members each of which bears some numerical value of a variable, e.g. of men measured according to height or of flowers classified according to numbers of petals. This variable we shall call a *variate.* If it can assume only a number of isolated values it will be called discontinuous, and if it can assume any value of a continuous range, continuous. The population of members will then correspond to a population of variate-values, and it is the properties of this latter population which we have to consider.

If the population consists of only a few members we can without much difficulty consider the population of variate-values exhibited by them ; but if, as usually happens, the aggregate is large (or, in a sense defined later, infinite), the set of variate-values has to be reduced in some way before the mind can grasp their significance. This is done by classification of the individuals into ranges of the variate. So far as possible the ranges should be equal, so that the numbers falling into different ranges are comparable. The interval is called the class-interval (or simply the interval) and the number of members bearing a variate-value falling into a given class-interval is the class-frequency (or simply the *frequency*). The manner in which the class-frequencies are distributed over the class-intervals is called the frequency-distribution (or simply the distribution).

1.7. Tables 1.1 and 1.2 give some frequency-distributions of observed populations classified according to a single variate. Table 1.1 shows the 1567 Local Government Areas of England and Wales distributed according to the variate " birth-rate." Here, for example, there were 7 districts with a birth-rate of between 5·5 and 6·5 per thousand, and 271 with a birth-rate between 13·5 and 14·5 per thousand. The general nature of

TABLE 1.1

Showing the Number of Local Government Areas in England with Specified Birth-rates per Thousand of Population.

(Material from the Registrar-General's Statistical Review of England and Wales for 1933.)

Birth-rate.	Number of Districts with Birth-rate in Specified Range.	Birth-rate.	Number of Districts with Birth-rate in Specified Range.
1·5 and not exceeding 2·5	1	13·5 and not exceeding 14·5	271
2·5 „ „ 3·5	2	14·5 „ „ 15·5	190
3·5 „ „ 4·5	2	15·5 „ „ 16·5	127
4·5 „ „ 5·5	3	16·5 „ „ 17·5	89
5·5 „ „ 6·5	7	17·5 „ „ 18·5	78
6·5 „ „ 7·5	9	18·5 „ „ 19·5	37
7·5 „ „ 8·5	14	19·5 „ „ 20·5	21
8·5 „ „ 9·5	41	20·5 „ „ 21·5	17
9·5 „ „ 10·5	83	21·5 „ „ 22·5	4
10·5 „ „ 11·5	131	22·5 „ „ 23·5	4
11·5 „ „ 12·5	192	23·5 „ „ 24·5	2
12·5 „ „ 13·5	242	TOTAL	1567

TABLE 1.2

Showing the Numbers of Persons in the United Kingdom liable to Sur-tax and Super-tax in the Year beginning 5*th April* 1931, *classified according to the Magnitude of their Annual Income.*

(From the Statistical Abstract for the United Kingdom for the Years 1913 and 1919–32, Cmd. 4489.)

Annual Income (£000).	Number of Persons.	Estimated Frequency per £500 Interval.
2 and not exceeding 2·5	23,988	23,988
2·5 „ „ 3	15,781	15,781
3 „ „ 4	17,979	8,989
4 „ „ 5	9,755	4,877
5 „ „ 6	5,921	2,960
6 „ „ 7	3,729	1,864
7 „ „ 8	2,546	1,273
8 „ „ 10	3,193	798
10 „ „ 15	3,616	362
15 „ „ 20	1,328	133
20 „ „ 25	679	68
25 „ „ 30	378	38
30 „ „ 40	372	19
40 „ „ 50	192	10
50 „ „ 75	182	4
75 „ „ 100	57	1
100 and over	94	?
Total number of persons	89,790	—

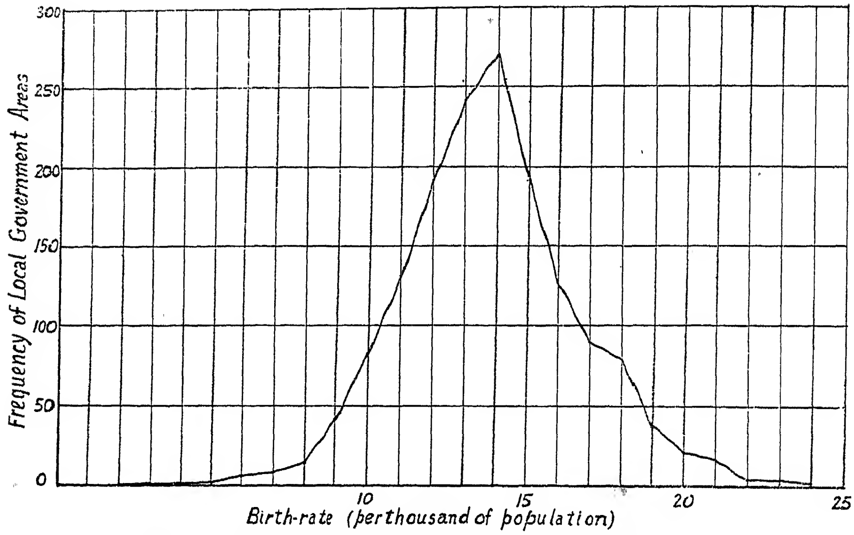


Fig. 1.1. Frequency Polygon of the Data of Table 1.1.

the distribution is shown in this table in a way which would be quite impossible if each of the 1567 districts were shown separately. The greatest number of districts fall within the range 13·5–14·5 per thousand and the frequencies tail off on either side of this value. Table 1.2 shows the number of persons subject to sur-tax and super-tax in the United Kingdom in 1931 classified according to the variate "income." The class-intervals here are unequal—a typical defect of official figures—and in the last column of the table is a reduction of the class-frequencies to comparability, namely, to frequency per £500 within the class-interval concerned. Looking at this column we see that the maximum frequency per £500 in this case is at the beginning of the frequency-distribution.

1.8. The frequency-distribution may be represented graphically. Measuring the variate-value along the x-axis and frequency per class-interval along the y-axis, we erect at the abscissa corresponding to the centre of each class-interval an ordinate equal to the frequency per unit interval in that interval. The ends of these ordinates are joined by straight lines, one to the next. The diagram so obtained is called a Frequency Polygon. Fig 1.1. shows the frequency polygon for the data of Table 1.1.

As a variant of this procedure we may erect on the abscissa range corresponding to each class-interval a rectangle whose area is proportional to the frequency in that interval. A diagram constructed in this way is called a Histogram. Fig. 1.2 shows such a histogram for the data of Table 1.2. It is evident that the histogram is a more suitable form of representation when the class-intervals are unequal.

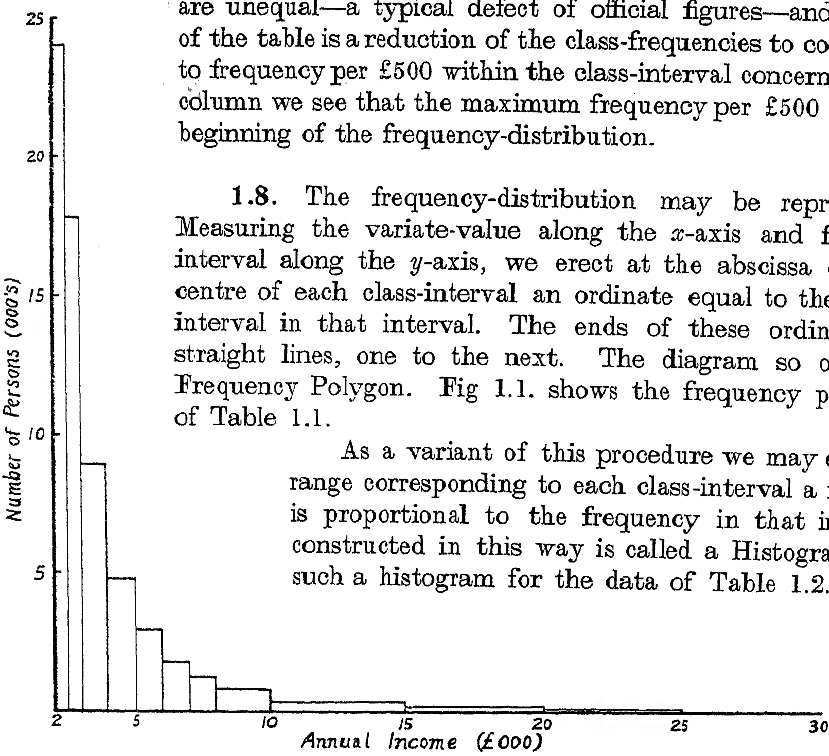


Fig. 1.2. Histogram of the Data of Table 1.2.

1.9. A few practical points in the tabulation of observed frequency-distributions may be noted.

(1) It has been remarked that wherever possible the class-intervals should be equal. The importance of this will be more appreciated in subsequent chapters; but it is already evident that comparability is difficult to carry out by inspection when there exist inequalities in class-intervals. On running the eye down the second column of Table 1.2, for example, we note that the frequencies in intervals 3–4 and 8–10 are greater than in the immediately preceding intervals; but this is merely due to a change in the width of the intervals at those points and, as is seen from the third column, the frequency per unit interval decreases steadily.

(2) It is important to specify the class-interval with precision. We not infrequently meet with such classifications as " 0–10, 10–20, 20–30," etc. To which interval is a member with variate-value 10 assigned? Obviously the classification is ambiguous if such values can in fact arise. We must either take the intervals " greater than or equal to 0 and less than 10, greater than or equal to 10 and less than 20," or make it clear what convention we use to allot a variate-value falling on the border between two neighbouring intervals, e.g. it might be decided to allot one-half of the member to each. There are various ways of indicating the class-interval in practical tables, e.g. " 10–, 20–, 30–" means " greater than or equal to 10 and less than 20," and so forth. Sometimes, where a continuous variate is concerned, there is an element of imprecision in the specification of the fineness to which the measurements are made; for example, if we are measuring lengths in centimetres to the nearest centimetre, an interval shown as " greater than 15 and less than 18" means an interval of " greater than 14.5 and less than 18.5." When the precision of the measurements is known we can specify an interval by its middle point, for example, in this case, 16.5.

TABLE 1.3

Showing the Number of Deaths from Scarlet Fever at Different Ages in England and Wales in 1933.

(Data from Registrar-General's Statistical Review of England and Wales for 1933, Tables Part I, Medical, supplemented by information supplied by him in correspondence.)

Age in Years.	Number of Deaths.	Number per Year.	Age in Years.	Number of Deaths.	Number per Year.
0–	16	16	40–	10	2·0
1–	69	69	45–	6	1·2
2–	89 } 322	89	50–	7	1·4
3–	74	74	55–	5	1·0
4–	74	74	60–	—	—
5–	213	42·6	65–	1	0·2
10–	70	14·0	70–	1	0·2
15–	27	5·4	75–	1	0·2
20–	26	5·2	80–	—	—
25–	17	3·4			
30–	12	2·4			
35–	11	2·2	TOTAL	729	—

(3) Remark (1) about the importance of equality of class-intervals should not be held to preclude the specification of frequencies in finer intervals where the frequency is changing very rapidly. Table 1.3, for instance, shows the number of deaths from scarlet fever in England and Wales in 1933 according to the variate "age at death." If the frequencies in the interval "0 and less than 5" were not subdivided and were thus shown as a total 322 for the interval, we might draw the conclusion from the uniformly decreasing number of deaths as the variate increases that the greatest number of deaths occurred in the first year of life. This is not so, as is shown by the individual frequencies in the first five years.

(4) Perhaps it is hardly necessary to add that the histogram is not a suitable method of representing data classified according to discontinuous variates. It shows the class-frequency uniformly dispersed over the whole interval, whereas if the variate is discontinuous, frequencies must necessarily be concentrated at certain points.

Frequency-Distributions : Discontinuous Variates

1.10. It will be useful at this stage to give some examples of the frequency-distributions which occur in practice.

Table 1.4 shows the distribution of digits in numbers taken from a four-figure telephone directory. The numbers were chosen by opening the directory haphazardly and taking the last two digits of all the numbers on the page except those in heavy type. The distribution is irregular, but from a cursory inspection of the table we are inclined to suppose that the digits occur approximately equally frequently in the larger population from which these 10,000 members were chosen. We shall see later (p. 193) that the divergences from the average frequency per digit, 1000, are not accidental sampling effects ; but at this stage it is sufficient to note that the data suggest for consideration a population of equally frequent members.

TABLE 1.4

Showing Number of Different Digits chosen haphazardly from the London Telephone Directory.

(M. G. Kendall and B. Babington Smith (1938), *Jour. Roy. Statist. Soc.*, **101**, 147.)

Digit . . .	0	1	2	3	4	5	6	7	8	9	TOTAL.
Frequency .	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Table 1.5 shows the distribution of a number of seed capsules of Shirley poppies according to the variate "number of stigmatic rays." The distribution in this case is

more regular, there being a maximum frequency at 13 and a steady decrease on either side.

TABLE 1.5

Showing the Frequencies of Seed Capsules on certain Shirley Poppies with Different Numbers of Stigmatic Rays.

(Cited from G. Udny Yule (1902), *Biometrika*, **2,** 89.)

Number of Stigmatic Rays.	Number of Capsules with said Number of Stigmatic Rays.	Number of Stigmatic Rays.	Number of Capsules with said Number of Stigmatic Rays.
6	3	14	302
7	11	15	234
8	38	16	128
9	106	17	50
10	152	18	19
11	238	19	3
12	305	20	1
13	315		
		TOTAL	1905

In Table 1.6, on the other hand, showing suicides among women in some German states in certain years according to the variate "number of suicides per year," the distribution reaches its maximum frequency in the region 1–3 suicides and then tails off rather slowly.

TABLE 1.6

Showing Suicides of Women in Eight German States in Fourteen years.

(Von Bortkiewicz, *Das Gesetz der kleinen Zahlen*, 1898.)

Number of Suicides	0	1	2	3	4	5	6	7	8	9	10 and over	TOTAL.
Frequency . . .	9	19	17	20	15	11	8	2	3	5	3	112

Frequency-Distributions : Continuous Variates

1.11. Table 1.7 shows a number of adult males in the United Kingdom (including, at the time of the collection of the data, the whole of Ireland), distributed according to the variate "height in inches." The frequency polygon is shown in Fig. 1.3. It will be seen that the distribution is almost symmetrical, there being a maximum ordinate at 67– inches and a steady decrease in frequency on either side of the maximum.

TABLE 1.7

Showing the Frequency-distributions of Statures for Adult Males born in the United Kingdom (including the whole of Ireland).

(Final Report of the Anthropometric Committee to the British Association, 1883, p. 256.)

As Measurements are stated to have been taken to the nearest $\frac{1}{8}$th of an inch, the class-intervals are here presumably $56\frac{15}{16}$–$57\frac{15}{16}$, $57\frac{15}{16}$–$58\frac{15}{16}$, and so on.

Height without Shoes (inches).	Number of Men within said Limits of Height.	Height without Shoes (inches).	Number of Men within said Limits of Height.
57–	2	69–	1063
58–	4	70–	646
59–	14	71–	392
60–	41	72–	202
61–	83	73–	79
62–	169	74–	32
63–	394	75–	16
64–	669	76–	5
65–	990	77–	2
66–	1223		
67–	1329		
68–	1230	TOTAL	8585

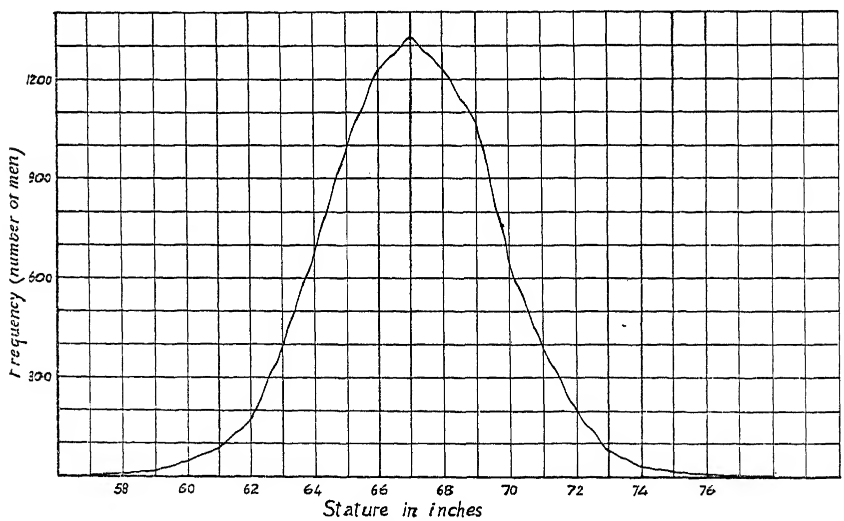


FIG. 1.3. Frequency-distribution of the Data of Table 1.7.

This more-or-less uniform "tailing off" of frequencies is very common in observed distributions, but the symmetrical property is comparatively rare. Table 1.1 is roughly symmetrical, but Tables 1.8 and 1.9, showing respectively a number of Australian marriages distributed according to bridegroom's age, and a number of dairy farms distributed according to costs of production of milk, illustrate that various degrees of asymmetry can occur. An extreme form is shown in Table 1.3.

TABLE 1.8

Showing Numbers of Marriages contracted in Australia, 1907–14, arranged according to the Age of Bridegroom in 3-Year Groups.

(From S. J. Pretorius (1930), *Biometrika*, **22**, 210.)

Age of Bridegroom (Central Value of 3-year Range, in years).	Number of Marriages.	Age of Bridegroom (Central Value of 3-year Range, in years).	Number of Marriages.
16·5	294	55·5	1,655
19·5	10,995	58·5	1,100
22·5	61,001	61·5	810
25·5	73,054	64·5	649
28·5	56,501	67·5	487
31·5	33,478	70·5	326
34·5	20,569	73·5	211
37·5	14,281	76·5	119
40·5	9,320	79·5	73
43·5	6,236	82·5	27
46·5	4,770	85·5	14
49·5	3,620	88·5	5
52·5	2,190		
		TOTAL	301,785

TABLE 1.9

Showing Numbers of Dairy Farms in England and Wales according to Cost of Production of Milk in 1935–6.

(Data from *Costs of Milk Production in England and Wales, Interim Report No. 2*, Agricultural Economics Research Institute, Oxford.)

Cost of Production (pence per gallon).	No. of Farms.	Cost of Production (pence per gallon).	No. of Farms.
4–	4	10–	65
5–	9	11–	40
6–	34	12–	15
7–	77	13–	4
8–	94	14–	5
9–	88	15–	2
		TOTAL	437

In this connection Table 1.10, showing a number of men distributed according to weight, is of interest for comparison with the height data of Table 1.7. The latter is symmetrical but the former is not.

TABLE 1.10

Frequency-distribution of Weights for Adult Males born in the United Kingdom.

(*Loc. cit.*, Table 1.7. Weights were taken to the nearest pound, consequently the true class-intervals are 89·5–99·5, 99·5–109·5, etc.)

Weight in lbs.	Frequency.	Weight in lbs.	Frequency.
90–	2	190–	263
100–	34	200–	107
110–	152	210–	85
120–	390	220–	41
130–	867	230–	16
140–	1623	240–	11
150–	1559	250–	8
160–	1326	260–	1
170–	787	270–	—
180–	476	280–	1
		TOTAL	7749

1.12. When the asymmetry of a distribution such as that of Table 1.3 becomes extreme we may be unable to determine whether, near the maximum ordinate, there is a fall on either side, or whether the maximum occurs right at the start of the distribution. This would have been the case in Table 1.3 if we had not the finer grouping for the first five years of life; and it is the case in Table 1.2, in which the maximum frequency apparently occurs at or very close to an income of £2,000 per annum. Asymmetrical distributions are sometimes called "skew"; and those such as Table 1.2 are called "J-shaped."

1.13. In rare cases the distribution may have maxima at both ends, as in Table 1.11,

TABLE 1.11

Showing the Frequencies of Estimated Intensities of Cloudiness at Greenwich during the Years 1890–1904 (excluding 1901) for the Month of July.

(Data from Gertrude E. Pearse (1928), *Biometrika*, 20A, 336.)

Degrees of Cloudiness.	Frequency.	Degrees of Cloudiness.	Frequency.
10	676	4	45
9	148	3	68
8	90	2	74
7	65	1	129
6	55	0	320
5	45		
		TOTAL	1715

showing a number of days distributed according to degree of cloudiness. This is known as a U-shaped distribution.

1.14. Distributions also occur which in general appearance resemble sections of the types already mentioned. A J-shaped distribution, for example, resembles the "tail" of the symmetrical distribution of Table 1.7. The suicide data of Table 1.6 may be regarded as a symmetrical distribution truncated just below the maximum ordinate by the impossibility of the occurrence of negative values of the variate. This sort of conception is sometimes useful in fitting curves to observed data—a given analytical curve may fit the data quite well in a certain variate range, but may also extend into regions where the data cannot, so to speak, follow it.

1.15. The distributions considered up to this point have one thing in common—they have only one maximum or, in the case of the U-shaped curve, only one minimum. Distributions also occur showing several maxima, Tables 1.12 and 1.13 being instances in point. The first, showing a number of deaths according to age at death, is typical of death distributions. Near the start of the distribution there is a maximum and a rapid fall in the frequency; there is an indication of another maximum about the age 20–25; and a pronounced maximum about the age 70–75, the frequencies beyond that point tailing off to zero. It is natural to wonder whether such a distribution can be usefully considered as three superposed distributions, a J-shaped distribution indicative of infantile mortality, a more or less symmetrical single-humped distribution with a maximum at 20–25, indicative of deaths at the adventurous age, and a skew distribution with a maximum at 70–75, the ordinary death curve of senescence.

TABLE 1.12

Showing the Number of Male Deaths in England and Wales for 1930–32, classified by Ages at Death.

(Data from Registrar-General's Statistical Review of England and Wales, 1933, text.

Age at Death (years).	Number of Deaths.	Age at Death (years).	Number of Deaths.
0–	97,290	55–	56,639
5–	11,532	60–	68,103
10–	7,305	65–	80,690
15–	13,062	70–	84,041
20–	16,741	75–	72,180
25–	16,126	80–	45,094
30–	15,673	85–	19,913
35–	18,345	90–	5,145
40–	23,778	95–	767
45–	33,158	100 and over	48
50–	43,821		
		TOTAL	729,442

TABLE 1.13

Showing Number of Trypanosomes from Glossina morsitans *classified according to Length in Microns.*

(From K. Pearson (1914–15), *Biometrika*, **10**, 112. Length presumably to nearest micron.)

Length (microns).	Frequency.	Length (microns).	Frequency.
15	7	26	110
16	31	27	127
17	148	28	133
18	230	29	113
19	326	30	96
20	252	31	54
21	237	32	44
22	184	33	11
23	143	34	7
24	115	35	2
25	130		
		TOTAL	2500

A similar dissection of a complex distribution could be undertaken for the data of Table 1.13, showing a number of trypanosomes from the tsetse fly, *Glossina morsitans*, classified according to length. We are led to suspect here that the distribution is composed of the addition of several others (and this, by the way, has led to a suggestion that the trypanosomes are a mixture of distinct types).

Frequency Functions and Distribution Functions

1.16. The examples given above illustrate the remarkable fact that the majority of the frequency-distributions encountered in practice possess a high degree of regularity. The form of the frequency polygons and histograms above suggests, almost inevitably, that our data are approximations to distributions which can be specified by smooth curves and simple mathematical expressions. This approach to the concept of the frequency *function*, however, requires some care, particularly for continuous distributions.

Consider in the first place a discontinuous distribution such as that of Table 1.4. Let us represent our variate by x. Then we may say that x can take any of the ten values $0, 1, \ldots 9$ and that the frequency of x, say $f(x)$, is given by the table, that is to say, $f(0) = 1026$, $f(1) = 1107$, $f(2) = 997$, and so on. The frequency table, in fact, defines the frequency function. Furthermore, most of the frequencies in the table are approximately 1,000, and we may then consider the observed distribution as approximating to that defined by

$$f(x) = 1000, \ x = 0, 1, \ldots 9 \qquad . \quad . \quad . \quad . \quad (1.1)$$

or, more generally, to the distribution

$$f(x) = k, \ x = 0, 1, \ldots 9 \qquad . \quad . \quad . \quad . \quad (1.2)$$

This is perhaps the simplest case of a discontinuous frequency function, $f(x)$ being a constant for all permissible values of x.

In Table 1.5 we have a discontinuous variate which can, theoretically, take an infinite number of values, namely, any one of the positive integers. In practice, of course, there must be a limit to the number of stigmatic rays which a poppy can possess, but since we do not know that limit we may imagine our variate as infinite in range. The frequency function for the table itself is again simply defined by the frequencies therein; but if we wish to proceed to a conceptual generalisation of such a table we must admit a discontinuous function $f(x)$ defined for all positive integral values of x. This occasions no difficulty provided that we are able to attach some meaning to the total frequency, i.e. that $\sum_{j=1}^{\infty} f(x_j)$ converges.

1.17. Consider now the case of a continuous variate. In the ordinary data of experience our distributions are invariably discontinuous, because our measurements can only attain a certain degree of accuracy. For instance, we are accustomed to suppose that the height of a man may in reality be any real number of inches in a certain range, say 50 to 80, such as 20π. In fact, we can measure heights only to a certain accuracy, say to the nearest thousandth of an inch. Our measurements thus consist of whole numbers (of thousandths) from 50,000 to 80,000, and such a number as 62,831·85 ($= 20\pi$ approximately) cannot appear. All physical measurements are subject to this limitation, but we accept it and nevertheless speak of our variables as "continuous," the underlying supposition being that the measurements are approximations to numbers which can fall anywhere in the arithmetic continuum.

1.18. With this understanding we can consider the distribution of grouped frequencies as leading to the concept of a frequency function for a continuous variate. If, in one of the distributions above, say that of Table 1.7, we were to subdivide the intervals, we should probably find that *up to a point* the resulting frequencies were smoother and smoother. The reader can verify the appearance of this effect for himself by grouping the data of Table 1.7 in intervals of 8, 4, and 2 inches. We cannot, however, take the process too far, because, with a finite population, continued subdivision of the interval would sooner or later result in irregular frequencies, there being only a few members in each interval. But we may suppose that for ranges Δx, not too small, the distribution may be specified by a function $f(x)\,\Delta x$, expressing that in the range $\pm \frac{1}{2}\Delta x$ centred at x the frequency is $f(x)\,\Delta x$, *wherever x may be in the permissible range of the variate.* We may suppose further that as Δx tends to zero the population is perpetually replenished so as to prevent the occurrence of small and irregular frequencies; and in this way we arrive at the concept of the frequency function for a continuous variable. We write

$$dF = f(x)\,dx \qquad . \quad . \quad . \quad . \quad . \quad . \quad (1.3)$$

expressing that the element of frequency dF between $x - \frac{1}{2}dx$ and $x + \frac{1}{2}dx$ is $f(x)\,dx$, for all x and for dx, however small.

1.19. This admittedly somewhat intuitive approach to the concept of the continuous frequency-distribution appears to be the best for statistical purposes, and is certainly the way in which the concept was originally reached. In formulating the axioms and postulates of a rigorous mathematical theory, however, the mathematician considers a rather more general function. There is as yet no thorough formulation of the theory required in this connection, and it would be alien to the primary purpose of this book to

attempt one, even if the space were available. We will merely indicate in broad outline the general approach.

1.20. We consider a function F which is defined at every point in a continuous range and is continuous, except perhaps at a denumerable number of points. We require that F shall be zero at the lower point of the range (which may be $-\infty$) and a constant N at the upper point (which may be $+\infty$) and that it shall not decrease at any point. Such a function is called a Distribution Function. It corresponds to the cumulated frequency of a frequency-distribution, N being the total frequency; for example, in Table 1.4, $F(x) = 0$ for $x < 0$, $F(x) = 1026$ for $x = 1$ and $x < 2$, $F(x) = 2133$ $(= 1026 + 1107)$ for $x = 2$ and $x < 3$, and so on. Here there are ten points of discontinuity for $F(x)$. These points are called "saltuses" (jumps) and $F(x)$ in this case is called a Step Function.

If there is no saltus in the range, $F(x)$ is continuous and monotonically increasing. If it possesses a derivative we have the equation in differentials

$$\begin{aligned} dF &= F'(x)\,dx \\ &= f(x)\,dx \end{aligned} \qquad (1.4)$$

corresponding to (1.3). $f(x)$ is called the Frequency Function. The mathematics of this branch of the subject is then that of the study of functions of the class $F(x)$ and $f(x)$.

1.21. The functions as thus defined are more general than those arrived at from the statistical approach in two ways: (i) $F(x)$ can increase monotonically in part of the range and then possess a saltus, i.e. the frequency may be continuous for a time and then suddenly discontinuous—in statistical practice a variate is either continuous or discontinuous, never both in different parts of the range; (ii) where no saltus exists $F(x)$ can exist without there existing a frequency function, just as a continuous function need not necessarily possess a derivative. In all the cases we shall consider, the existence of a continuous variate will be accompanied by the existence of a frequency function.

The function $F(x)$ is sometimes called a Probability Function, for reasons which will become evident in Chapter 7 when we consider the theory of probability. Essentially, however, it has nothing to do with probability and we shall use the term "distribution function" only.

1.22. If the discontinuous frequency function is $f(x)$, and $F(x)$ is taken to be the total frequency less than *or equal to* x, we have

$$F(x_r) = \sum_{j=1}^{r} f(x_j) \qquad (1.5)$$

In the continuous case

$$\begin{aligned} F(x) &= \int_a^x dF \\ &= \int_a^x f(x)\,dx \end{aligned} \qquad (1.6)$$

where the range is a to b. We now introduce two conventions which simplify these expressions to some extent. We shall suppose, unless the contrary is specified, that in these mathematical expressions our frequencies are always expressed as proportions of the total frequencies, so that the total frequency is unity and the sum or integral over the whole range of the frequency function is also unity, i.e. $F(b) = 1$. Secondly, to avoid the

constant specification of the limits a and b we may, without loss of generality, suppose that $F(x)$ and $f(x)$ are zero for any x less than a, and that $F(x) = 1$ and $f(x) = 0$ for any x greater than b. With this convention we may write

$$\begin{aligned} F(x_r) &= \sum_{j=-\infty}^{r} f(x_j) \\ F(x) &= \int_{-\infty}^{x} f(x)\,dx \end{aligned} \tag{1.7}$$

and

$$\begin{aligned} \sum_{j=-\infty}^{\infty} f(x_j) &= F(\infty) - F(-\infty) = 1 \\ \int_{-\infty}^{\infty} f(x)\,dx &= F(\infty) - F(-\infty) = 1 \end{aligned} \tag{1.8}$$

Where it is necessary to take account of the total frequency N we may do so by multiplying by N frequencies given by the frequency function. In our convention $F(x)$ is always continuous on the left.

Excursus on Stieltjes Integrals

1.23. The distinction between discontinuous and continuous distributions, though real and important for statistical purposes, is something of a nuisance in mathematical investigations, and to avoid the necessity of stating all our theorems twice we shall use a type of integral due to Stieltjes. In effect, this integral subsumes under one summatory process the finite summation denoted by Σ and the ordinary integral denoted by $\int$.

Suppose, in fact, that $F(x)$ is a distribution function as we have defined it. Let $\psi(x)$ be a continuous function in the range of $F(x)$, which we will take in the first instance to be finite, a to b. Divide the range into n intervals at points $a = x_0, x_1, x_2, \ldots x_{n-1}, x_n = b$. Take ξ_0 in the range a to x_1, ξ_1 in the range x_1 to x_2, and so on. Let

$$\begin{aligned} S = \psi(\xi_0)\{F(x_1) - F(a)\} + \psi(\xi_2)\{F(x_2) - F(x_1)\} \\ + \ldots + \psi(\xi_n)\{F(b) - F(x_{n-1})\} \end{aligned} \tag{1.9}$$

It may be shown that as the size of the intervals $x_{r+1} - x_r$ tends to zero uniformly, S tends to a limit which is independent of the location of the points ξ or of the boundary points of the intervals. We then write this limit

$$\int_{c}^{l} \psi(x)\,dF \tag{1.10}$$

and define it as the Stieltjes integral of $\psi(x)$ with respect to $F(x)$.

As for the case of ordinary integrals, we may now consider a and b as tending to infinity and write, for example

$$\int_{-}^{x} \psi(x)\,dF,$$

provided that the limit exists.

In particular, if $\psi(x) = 1$, we have the distribution function

$$F(x) = \int_{-c}^{x} dF.$$

1.24. If $F(x)$ is the distribution function of a distribution possessing a continuous frequency function, the Stieltjes integral becomes the ordinary integral

$$\int_a^x \psi(x) f(x)\, dx,$$

and thus includes ordinary integration as a particular case. If $F(x)$ is the distribution function of a discontinuous distribution, that is to say, is a step function, a term such as $F(x_{r+1}) - F(x_r)$ will vanish unless there is a saltus in the range x_r to x_{r+1}. The sum S of (1.9) must then tend to the limit (since it does tend to a limit) $\Sigma\, \psi(x_r)\, f(x_r)$, i.e. to the ordinary summation of a series. The Stieltjes integral thus also includes such summation as a particular case.

1.25. Many of the theorems of ordinary integration are true of the Stieltjes integral. We shall frequently require the following:

$$\left|\int_a^b \psi\, dF\right| \leqslant \int_a^b |\,\psi\,|\, dF \qquad . \quad (1.11)$$

$$\leqslant M\int_a^b dF$$

$$\leqslant M \qquad . \quad (1.12)$$

where M is the upper bound of $\psi(x)$ in the range (a, b).

$$\int_a^b \psi\, dF = \psi(\xi)\int_a^b dF \qquad . \quad (1.13)$$

where ξ is a value of x in the range (a, b).

If a and b are finite

$$\int_a^b \sum_{j=1}^{\infty} f_j(x)\, dF = \sum_{j=1}^{\infty}\int_a^b f_j(x)\, dF, \qquad . \quad (1.14)$$

provided that $\Sigma f_j(x)$ converges uniformly in the range. The theorem is not necessarily true if a and b, or one of them, are infinite.

The ordinary rules of partial integration are also applicable to Stieltjes integrals.

Variate Transformations

1.26. Suppose we have a new variate ξ related to x by some functional equation

$$x = x(\xi), \qquad . \quad . \quad . \quad . \quad . \quad . \quad (1.15)$$

ξ being continuous and differentiable in x throughout the range of x, and vice-versa. We have then the equation in differentials

$$dx = x'\, d\xi = \frac{dx}{d\xi}\, d\xi. \qquad . \quad (1.16)$$

Consequently, for a continuous distribution

$$F(x) = \int_{-\infty}^{x} dF = \int_{-\infty}^{x} f(x)\, dx$$

$$= \int_{\infty} f(x)\frac{dx}{d\xi}\, d\xi,$$

and consequently we may write the distribution as

$$dF = f\{x(\xi)\}\frac{dx}{d\xi}\,d\xi \qquad . \qquad (1.17)$$

expressing that an element of frequency between $\xi - \frac{1}{2}\,d\xi$ and $\xi + \frac{1}{2}\,d\xi$ is $f\{x(\xi)\}\,\frac{dx}{d\xi}$. The equation determining the frequency function may then be transformed as if it were an equation in differentials. Such transformations are important in the theory of continuous distributions. By their means many mathematically specified distributions may be reduced to known forms, either exactly or approximately.

For example, a distribution which we shall have to study in the theory of sampling is

$$dF = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{\chi^2}{2}}\chi^{\nu-1}\,d\chi \qquad 0 \leqslant \chi \leqslant \infty.$$

It is readily verified by integration that $F(\infty) = 1$.

By the transformation $\frac{\chi^2}{2} = t$ we reduce this to

$$dF = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} e^{-t}\,t^{\frac{\nu}{2}-1}\,dt \qquad 0 \leqslant t \leqslant \infty$$

a well-known form in analysis, the distribution function being the incomplete Γ-function

$$F(t) = \int_0^t \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} e^{-t}\,t^{\frac{\nu}{2}-1}\,dt$$

$$= \Gamma_t\left(\frac{\nu}{2}\right)\Big/\Gamma\left(\frac{\nu}{2}\right).$$

Again, the distribution

$$dF = \frac{y_0}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}\,dt \qquad -\infty \leqslant t \leqslant \infty$$

(y_0 being chosen so that $F(\infty) = 1$), a symmetrical peaked distribution of infinite range rather like that of Fig. 1.3, may, by the substitution of $t = \sqrt{\nu}\tan\theta$, be transformed into

$$dF = \frac{y_0\sqrt{\nu}\sec^2\theta\,d\theta}{\sec^{\nu+1}\theta} \qquad -\frac{\pi}{2} \leqslant \theta \leqslant \frac{\pi}{2}$$

$$= y_0\sqrt{\nu}\cos^{\nu-1}\theta\,d\theta,$$

a distribution of finite range $-\frac{\pi}{2}$ to $+\frac{\pi}{2}$, but still symmetrical. Putting now $\sin\theta = \xi$, we have

$$dF = y_0\sqrt{\nu}(1 - \xi^2)^{\frac{\nu-2}{2}}\,d\xi \qquad 1 \leqslant \xi \leqslant 1,$$

and again $\xi^2 = x$,

$$dF = y_0\sqrt{\nu}(1 - x)^{\frac{\nu-2}{2}}x^{-\frac{1}{2}}\,dx \qquad 0 \leqslant x \leqslant 1.$$

The effect on the range of this last substitution is to be noted. ξ ranges from -1 to $+1$, and as it does so x ranges from $+1$ to 0 and back to $+1$. The distribution function of the x-distribution, $F(x)$ from 0 to x, is thus that of the ξ-distribution from $-\xi^2$ to ξ^2. Whenever substitutions are made under which there is not a $(1, 1)$ continuous relation between the variates, points such as this require some watching.

1.27. There is one variate transformation which is worth special attention. In the distribution

$$dF = f(x)\, dx$$

put

$$\xi = \int_{-\infty}^{x} f(x)\, dx.$$

Then

$$dF = f(x) \frac{dx}{d\xi} d\xi$$

$$= \frac{f(x)}{f(x)} d\xi$$

$$= d\xi \qquad 0 \leqslant \xi \leqslant 1 \qquad . \quad (1.18)$$

so that the distribution is transformed into the very simple "rectangular" form in which all values of the variate from 0 to 1 are equally frequent. Any continuous distribution can be transformed into the rectangular form; and it follows that there exists at least one transformation which will transform any continuous frequency-distribution into any other continuous frequency-distribution, viz. the transformation which transforms one into the rectangular form coupled with the reverse of that which transforms the other into the rectangular form.

The Genesis of Frequency-Distributions

1.28. Up to this point we have not inquired into the origin of the various observed frequency-distributions which have been adduced in illustration. Certain of them may be considered apart from any question of origination from a larger population. The death distribution of Table 1.12 is an example; if we are interested only in the distribution of male deaths in England and Wales in 1930–32 the whole of the population under consideration is before us.

But in the great majority of cases the population which we are able to examine is only part of a larger population on which our main interest is centred. The height distribution of Table 1.7 is only a part of the population of men in the United Kingdom living at the time of the inquiry, and it is mainly of importance in the light of the information which it gives us about that population. Similarly the distribution of farms of Table 1.9 is largely of interest in the information it gives about costs of milk production for the whole country.

1.29. In the two cases just mentioned, height and costs of milk production, we have information about a certain sample of individuals chosen from an existing population. Only lack of time and opportunity prevents us from examining the whole population. It sometimes happens, however, that we have data which do not emanate from a finite existent population in this way. Table 1.14 is an example. It shows the distribution of throws with dice.

TABLE 1.14

Showing the Number of Successes (throws of 4, 5 or 6) with Throws of 12 Dice.
(Weldon's data, cited by F. Y. Edgeworth, *Encyclopædia Britannica*, 11th ed., **22**, 39.)

Number of Successes.	Frequency.	Number of Successes.	Frequency.
0	0	7	847
1	7	8	536
2	60	9	257
3	198	10	71
4	430	11	11
5	731	12	0
6	948		
		TOTAL	4096

Now it is clear that, in a sense, we have not in these data got a complete population, for we can add to them by further casting of the dice. But these further throws do not exist in the sense that the unexamined men of the United Kingdom or the unexamined dairy farms of England and Wales exist. They have a kind of hypothetical existence conferred on them by our notion of the throwing of the dice.

Even distributions which appear at first sight to be existent may be considered in this light. The trypanosome distribution of Table 1.13, for instance, was obtained from certain tsetse flies. We may consider it as a sample of all the tsetse flies in existence, whether harbouring trypanosomes or not—an existent population; but we may also consider it as a sample of what the distribution would be if all the tsetse flies were infected with trypanosomes—a hypothetical population.

The population conceived of as parental to an observed distribution is fundamental to statistical inference. We shall take up this matter again in later chapters when we consider the sampling problem. The point is mentioned here because it will occasionally arise before we reach that chapter. It must be emphasised that the distinction between existent and hypothetical universes is not merely a matter of ontological speculation—if it were we could safely ignore it—but one of practical importance when inferences are drawn about a population from a sample generated from it.

Multivariate Distributions

1.30. In the foregoing sections we have considered the members of a population according to a single variate, and the frequency-distributions may thus be called univariate. The work may be readily generalised to include populations of members considered according to two or more variates, yielding bivariate, trivariate . . . multivariate frequency distributions. Table 1.15, for example, shows the distribution of a number of beans according to both length and breadth. The border frequencies show the univariate distributions of the beans according to length and breadth separately, and the body of the table shows how the two qualities vary together.

TABLE 1.15

Showing Frequencies of Beans with specified Lengths and Breadths.
(Johannsen's data, cited by S. J. Pretorius (1930), *Biometrika*, **22**, 110.)

Lengths in millimetres (central values).

	17	16·5	16	15·5	15	14·5	14	13·5	13	12·5	12	11·5	11	10·5	10	9·5	TOTALS.
9·125	—	2	—	—	3	—	—	—	—	—	—	—	—	—	—	—	5
8·875	4	8	17	19	—	—	—	—	—	—	—	—	—	—	—	—	48
8·625	2	23	101	156	93	23	2	—	—	—	—	—	—	—	—	—	400
8·375	—	18	105	494	574	227	56	9	—	—	—	—	—	—	—	—	1483
8·125	—	4	44	375	956	913	362	73	12	3	—	—	—	—	—	—	2742
7·875	—	—	7	81	385	871	794	330	89	19	3	—	—	—	—	—	2579
7·625	—	—	1	4	65	236	469	361	175	55	27	4	—	—	—	—	1397
7·375	—	—	—	—	6	23	91	137	124	78	37	22	11	—	1	—	530
7·125	—	—	—	—	—	1	13	18	28	35	25	32	11	6	1	—	170
6·875	—	—	—	—	—	—	—	1	9	8	21	12	13	7	1	—	72
6·625	—	—	—	—	—	—	—	—	—	—	2	—	1	4	3	—	10
6·375	—	—	—	—	—	—	—	—	—	1	—	—	—	1	1	1	4
TOTALS	6	55	275	1129	2082	2294	1787	929	437	199	115	70	36	18	7	1	9440

As for the univariate case, the variates may be discontinuous or continuous and we sometimes meet cases in which one variate is of one kind and one of the other.

1.31. In generalisation of the frequency polygon and the histogram we may construct 3-dimensional figures to represent the trivariate distribution. Imagine a horizontal plane containing a pair of perpendicular axes and ruled like a chessboard into cells, the ruled lines being drawn at points corresponding to the terminal points of class-intervals. At

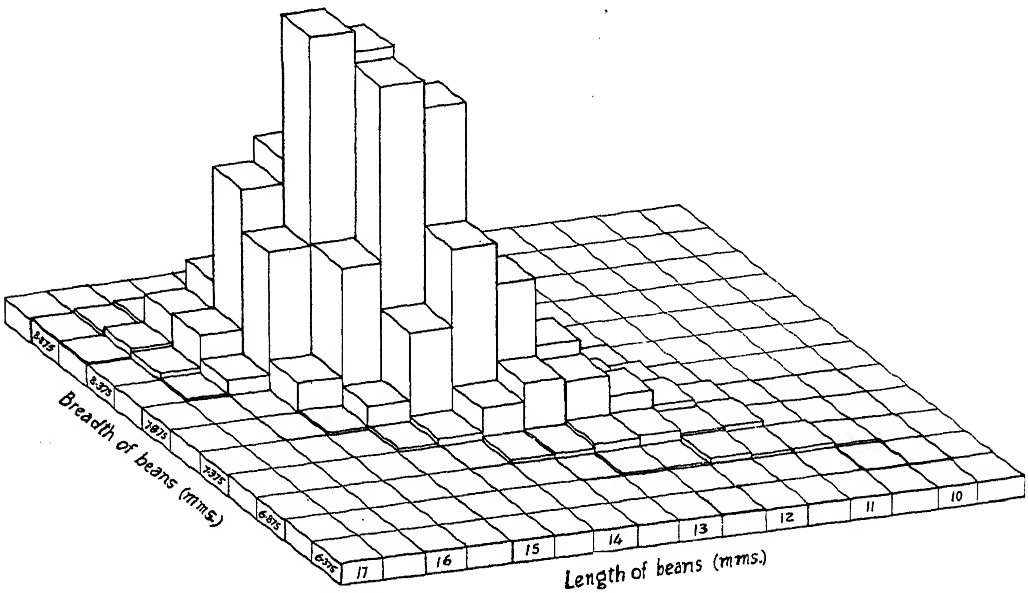


FIG. 1.4. Bivariate Histogram of the Data of Table 1.15.

the centre of each interval we erect a vertical line proportional in length to the frequency in that interval. The summits of these verticals are joined, each to the four summits of verticals in the neighbouring cells possessing the same values of one or the other variate. The resulting figure is the bivariate frequency polygon or Stereogram.

Similarly we may erect on each cell a pillar proportional in volume to the frequency in that cell and thus obtain a bivariate histogram. Fig. 1.4 shows such a figure for the bean data of Table 1.15.

1.32. We may write the bivariate distribution with variates x_1, x_2, as

$$dF = f(x_1, x_2)\, dx_1\, dx_2 \quad . \quad . \quad . \quad . \quad . \quad (1.19)$$

With the usual conventions we shall then have for the bivariate distribution function

$$F(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(x_1, x_2)\, dx_1\, dx_2 \quad . \quad (1.20)$$

this integral also being understood in the Stieltjes sense, reducing to ordinary integration if $f(x_1, x_2)$ is continuous and to ordinary summation if it is discontinuous.

Independence

1.33. If there are two distribution functions F_1, F_2, such that

$$F(x_1, x_2) = F_1(x_1)\, F_2(x_2) \quad . \quad . \quad . \quad . \quad . \quad (1.21)$$

then x_1 and x_2 are said to be independent. Where frequency functions exist we have

$$\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(x_1, x_2)\, dx_1\, dx_2 = \int_{-\infty}^{x_1} f_1(x_1)\, dx_1 \int_{-\infty}^{x_2} f_2(x_2)\, dx_2$$

giving

$$f(x_1, x_2) = f_1(x_1)\, f_2(x_2) \quad (1.22)$$

It is readily seen that this definition of statistical independence conforms to the colloquial use of the word and also to its mathematical use. The distribution of x_2 for any fixed x_1 (e.g. the distribution in a row or column of the bivariate frequency table) is the same whatever the fixed value of x_1, that is to say, the distribution of x_2 is independent of x_1.

Two variates which are not independent are said to be dependent. Evidently those of Table 1.15 are dependent, for the distributions in rows or in columns are far from similar.

Generally, n variates are independent if

$$F(x_1 \ldots x_n) = F_1(x_1) \ldots F_n(x_n).$$

1.34. Transformations of the variate for bi- and multivariate distributions follow the ordinary laws for the transformation of differentials. For example, if

$$dF = f(x_1, x_2)\, dx_1\, dx_2$$
$$x_1 = x_1(\xi_1, \xi_2) \qquad x_2 = x_2(\xi_1, \xi_2)$$

we have

$$dF = f\{x_1(\xi_1, \xi_2), x_2(\xi_1, \xi_2)\} J\, d\xi_1, d\xi_2 . \quad (1.23)$$

where J is the Jacobian

$$J = \frac{\partial(x_1, x_2)}{\partial(\xi_1, \xi_2} \quad \begin{vmatrix} \dfrac{\partial x_1}{\partial \xi_1} & \dfrac{\partial x_2}{\partial \xi_1} \\ \dfrac{\partial x_1}{\partial \xi_2} & \dfrac{\partial x_2}{\partial \xi_2} \end{vmatrix}$$

and is to be taken with a positive sign in (1.23).

Consider, for example, the distribution

$$dF = z_0 \exp\left\{-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\} dx_1 dx_2 \qquad -\infty \leqslant x_1, x_2 \leqslant \infty. \quad (1.24)$$

z_0, as usual, being chosen so that the total frequency is unity. The variates are evidently dependent.

Put

$$\xi_1 = \frac{x_1}{\sigma_1} - \frac{\rho x_2}{\sigma_2}$$

$$\xi_2 = (1-\rho^2)^{\frac{1}{2}} \frac{x_2}{\sigma_2}$$

We have

$$\frac{\partial(\xi_1, \xi_2)}{\partial(x_1, x_2)} \begin{vmatrix} \dfrac{1}{\sigma_1} & -\dfrac{\rho}{\sigma_2} \\ 0 & (1-\rho^2)^{\frac{1}{2}}.\dfrac{1}{\sigma_2} \end{vmatrix} = \frac{(1-\rho^2)^{\frac{1}{2}}}{\sigma_1\sigma_2}$$

and

$$\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2} = \xi_1^2 + \xi_2^2$$

The distribution then becomes

$$dF = \frac{z_0\sigma_1\sigma_2}{(1-\rho^2)^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}(\xi_1^2 + \xi_2^2)\} d\xi_1, d\xi_2 \qquad (1.25)$$

$$= \frac{z_0\sigma_1\sigma_2}{(1-\rho^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\xi_1^2} d\xi_1 \; e^{-\frac{1}{2}\xi_2^2} d\xi_2 \quad . \qquad . \qquad (1.26)$$

The transformed variates ξ_1 and ξ_2 are thus independent.

NOTES AND REFERENCES

The collection of definitions of statistics by Willcox (1935) has already been referred to in the text.

Examples of practical frequency-distributions will be found in most statistical journals, particularly *Biometrika*.

As to the mathematical basis of the theory of frequency-distributions, there appears to be no account in English. The reader who is interested should, however, make a point of reading two French works, that by Lévy and those by Fréchet in the Borel *Traité*. Both these are written from the standpoint of the theory of probability, but the basic ideas of the theory of frequency-distributions are the same whether probability is concerned or not.

Borel, E., *Traité du Calcul des Probabilités*, Gauthier-Villars, Paris. A series of brochures written under the general editorship of M. Borel. See particularly the two by M. Fréchet called "Nouveaux Recherches."

Lévy, P., *Calcul des Probabilités*, Gauthier-Villars, Paris.

Shohat, J. (1929), "Stieltjes Integrals in Mathematical Statistics," *Ann. Math. Statist.*, **1**, 73.

Willcox, W. F. (1935), "Definitions of Statistics," *Revue de l'Inst. Int. de Statistique*, **3**, 388.

EXERCISES

1.1. Draw frequency polygons or histograms of the following distributions :—

TABLE 1.16

Frequency-Distribution of Successes in Twelve Dice thrown 4096 Times, a Throw of 6 Points reckoned as a Success.

(Weldon's data ; *loc. cit.*, Table 1.14.)

Number of Successes . .	0	1	2	3	4	5	6	7 and over	TOTAL.
Number of Throws . .	447	1145	1181	796	380	115	24	8	4096

TABLE 1.17

Frequency-Distribution of Size of Firms in the Food, Drink and Tobacco Trades of Great Britain.

(Final Report of the Fourth Census of Production, 1930, Part III. The table shows the number of firms employing, on an average, certain numbers of persons.)

Size of Firm (Average Numbers Employed).	11–24	25–49	50–99	100–199	200–299	300–399	400–499	500–749	750–999	1000–1499	1500 and over	TOTAL.
Number of Firms .	2245	1449	771	439	164	75	36	54	31	23	29	5316

TABLE 1.18

Frequency-Distribution of Plots according to Yield of Grain in Pounds from Plots of $\frac{1}{500}$th Acre in a Wheat Field.

(Mercer and Hall (1911), *Jour. Agr. Science*, **4**, 107.)

Yield of Grain in pounds per $\frac{1}{500}$th Acre. (Central value of range).	2·8	3·0	3·2	3·4	3·6	3·8	4·0	4·2	4·4	4·6	4·8	5·0	5·2	TOTAL.
Number of Plots . . .	4	15	20	47	63	78	88	69	59	35	10	8	4	500

TABLE 1.19

The Percentages of Deaf-mutes among Children of Parents One of whom at least was a Deaf-mute, for Marriages producing Five Children or More.

(Compiled by G. Udny Yule from material in *Marriages of the Deaf in America*, ed. E. A. Fay, Volta Bureau, Washington, 1898. Where a family fell on the border line between two class-intervals one-half was assigned to each.)

Percentage of Deaf-mutes.	Number of Families.	Percentage of Deaf-mutes.	Number of Families.
0–20	220	60–80	5·5
20–40	20·5	80–100	15
40–60	12		
		TOTAL	273

TABLE 1.20

Showing the Frequency-Distribution of Fecundity, i.e. the Ratio of the Number of Yearling Foals produced to the Number of Coverings, for Brood-mares (Racehorses) covered Eight Times at least.

(Pearson, Lee and Moore (1899), *Phil. Trans.*, A, **192**, 303. Where a case fell on the border between two intervals, one-half was assigned to each.)

Fecundity.	Number of Mares with Fecundity between the Given Limits.	Fecundity.	Number of Mares with Fecundity between the Given Limits.
1/30– 3/30	2	17/30–19/30	315
3/30– 5/30	7·5	19/30–21/30	337
5/30– 7/30	11·5	21/30–23/30	293·5
7/30– 9/30	21·5	23/30–25/30	204
9/30–11/30	55	25/30–27/30	127
11/30–13/30	104·5	27/30–29/30	49
13/30–15/30	182	29/30–1	19
15/30–17/30	271·5		
		TOTAL	2000·0

TABLE 1.21

Showing Numbers of Sentences of given Lengths in Passages from Macaulay's Essays on Bacon and on Chatham.

(From G. Udny Yule (1939), *Biometrika*, **30**, 363.)

Length of Sentence in Words.	Number of Sentences.	Length of Sentence in Words.	Number of Sentences.
1–5	46	66–	2
6–	204	71–	4
11–	252	76–	8
16–	200	81–	2
21–	186	86–	2
26–	108	91–	1
31–	61	96–	2
36–	68	101–	1
41–	38	106–	—
46–	24	111–	1
51–	20	116–	—
56–	12	121–	1
61–	8		
		TOTAL	1251

TABLE 1.22

Showing the Numbers of Old Egyptian Skulls with Specified Lengths of the Left Occipital Bone in millimetres.

(From T. L. Woo (1930), *Biometrika*, **22**, 324.)

Length (central values).	Frequency.	Length (central values).	Frequency.
84·5	12	102·5	74
86·5	12	104·5	68
88·5	32	106·5	36
90·5	48	108·5	18
92·5	79	110·5	7
94·5	116	112·5	4
96·5	104	114·5	4
98·5	126	116·5	—
100·5	123	118·5	1
		TOTAL	864

TABLE 1.23

Showing the Number of Women Aborting at Specified Term in Weeks.

(From T. V. Pearce (1930), *Biometrika*, **22**, 250.)

Term (weeks).	Frequency.	Term (weeks).	Frequency.
4	3	17	13
5	7	18	14
6	10	19	8
7	13	20	4
8	14	21	2
9	29	22	10
10	22	23	4
11	21	24	4
12	18	25	3
13	28	26	4
14	16	27	6
15	19	28	1
16	10		
		TOTAL	283

1.2. Sketch the following curves and compare their shapes with those of the distributions in the previous exercise :—

$$y = y_0 e^{\quad 2} \qquad -\infty \leqslant x \leqslant \infty$$
$$y = y_0 e^{-x} \qquad 1 \leqslant x \leqslant \infty$$
$$y = y_0 x^{-\gamma} \qquad \gamma > 1,\ 0 \leqslant x \leqslant \infty$$
$$y = \frac{y_0}{(1+x^2)^n} \qquad n > 0,\ -\infty \leqslant x \leqslant$$
$$y = y_0(1-x)^a x^b \qquad a,\ b > 1,\ 0 \leqslant x \leqslant 1$$
$$y = y_0 e^{-x} x^{\gamma} \qquad \gamma > 1,\ 0 \leqslant x \leqslant \infty$$
$$y = y_0(1-x^2)^a \qquad a < 0,\ -1 \leqslant x \leqslant 1.$$

1.3. Show that the following distributions can all be transformed into the type

$$dF = y_0(1-x)^{p-1}x^{q-1}\,dx \qquad 0 \leqslant x \leqslant 1$$

and find the transformations :

$$dF = r_0(1-r^2)^{\frac{n-4}{2}}\,dr \qquad -1 \leqslant r \leqslant 1$$

$$dF = \frac{\quad}{\left(\ +\frac{t^2}{\nu}\right)^{\frac{\nu}{}}}\,dt \qquad -\infty \leqslant t \leqslant \infty$$

$$dF = \frac{z_0 e^{\nu_1}}{(\nu_1 e^{2z}+\nu_2)^{\frac{\nu_1+\nu_2}{2}}}\,dz \qquad -\infty \leqslant z \leqslant \infty$$

(All these distributions are important in statistical theory. The distribution to which they are reduced is called the Type I or B-distribution.)

1.4. Sketch the stereograms or bivariate histograms of the following distributions:

TABLE 1.24

Number of Families deficient in Room Space in 95 crowded London Wards.
(Census of 1931, *Housing Report*, p. xxxii.)

	Standard Room Requirement (Rooms).							
Families deficient by	2	3	4	5	6	7	8	TOTALS.
1 room	12,999	18,198	7,724	2,170	164	19	..	41,274
2 rooms	..	3,054	4,479	1,448	221	15	1	9,218
3 rooms	..	..	310	508	106	4	1	929
4 rooms	..	..	..	10	21	4	..	35
TOTALS	12,999	21,252	12,513	4,136	512	42	2	51,456

TABLE 1.25

Number of Cows Distributed according to (1) *Age in Years and* (2) *Yield of Milk per Week* in *4912 Ayrshire Cows.*
(Data from J. F. Tocher (1928), *Biometrika*, 20B, 106.)

(1) Age in Years.

(2) Yield of Milk per Week (Gallons). (Central Value of Interval.)	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	TOTALS.
8	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
9	—	2	2	—	1	—	—	—	—	—	—	—	—	—	—	—	5
10	3	5	1	1	3	—	—	—	—	—	—	—	—	—	—	—	13
11	2	10	8	7	1	—	1	—	2	1	—	1	—	—	—	—	33
12	2	25	17	9	5	4	4	2	1	1	—	—	—	1	—	—	71
13	9	76	29	18	9	2	4	1	1	1	—	1	—	—	—	—	151
14	11	76	57	38	23	9	7	6	4	2	3	—	—	—	—	—	236
15	11	115	79	43	34	24	11	8	4	5	1	2	1	—	1	—	339
16	15	149	119	74	59	23	23	16	9	7	4	—	—	—	1	—	499
17	16	148	131	94	58	34	32	15	12	6	5	—	1	—	—	—	552
18	11	146	132	83	73	49	39	22	17	6	5	1	1	—	—	—	585
19	10	117	112	113	87	51	35	33	11	10	2	3	1	—	—	1	586
20	8	97	107	79	69	51	25	30	13	10	3	3	—	—	1	—	496
21	3	63	93	88	70	49	31	29	9	7	4	—	1	—	1	—	448
22	5	42	63	49	45	32	14	18	10	3	1	2	—	—	—	—	284
23	1	19	33	38	38	27	17	17	12	7	1	2	2	—	—	—	214
24	2	20	23	34	27	19	13	9	3	2	1	—	—	—	—	—	153
25	3	10	15	22	17	20	8	10	3	4	—	—	—	—	—	—	112
26	—	7	13	7	4	15	2	4	2	3	1	—	—	—	—	—	58
27	—	2	7	9	5	5	4	2	—	—	—	—	—	1	—	—	35
28	—	—	2	1	4	2	1	1	2	—	—	—	—	—	—	—	13
29	—	—	2	2	4	1	3	—	3	—	—	—	—	—	—	—	15
30	—	—	—	—	—	2	—	—	2	—	—	—	—	—	—	—	4
31	—	—	2	1	—	—	2	—	—	—	—	—	—	—	—	—	5
32	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	2
33	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
34	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
TOTALS	112	1129	1047	812	636	419	276	223	122	75	32	15	7	2	4	1	4912

TABLE 1.26

Distribution of Weekly Returns according to (1) *Call Discount Rates and* (2) *Percentage of Reserves on Deposits in New York Associated Banks.*

(From *Statistical Studies in the New York Money Market*, by J. P. Norton. Publications of the Department of the Social Sciences, Yale University; The Macmillan Co., 1902.) Note that, after the column headed 8 per cent., blank columns have been omitted to save space.

(1) Call Discount Rates.

(2) Percentage Ratio of Reserves to Deposits.	1	1·5	2	2·5	3	3·5	4	4·5	5	5·5	6	6·5	7	7·5	8	9	10	12	15	20	25	Totals
21	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1	—	—	—	—	—	2
22	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
23	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
24	—	—	—	—	—	—	—	1	—	—	2	—	—	—	3	—	2	—	—	—	1	9
25	—	—	—	—	—	1	2	6	4	4	11	1	2	—	6	1	2	1	1	—	—	42
26	—	—	—	—	2	6	13	12	16	6	11	4	7	—	—	2	2	—	1	1	2	85
27	—	1	10	9	14	12	15	17	19	9	9	3	4	—	1	—	—	—	1	—	—	124
28	—	5	30	23	20	11	7	3	7	1	2	2	3	—	—	—	1	—	—	—	—	115
29	3	9	48	17	16	3	6	3	1	—	—	—	2	—	—	—	—	—	—	—	1	109
30	1	12	12	10	8	4	4	—	2	—	—	—	—	—	—	—	—	—	—	—	—	53
31	8	10	6	2	4	2	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	36
32	15	14	10	8	5	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	53
33	15	8	4	1	—	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	32
34	2	11	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
35	8	5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
36	7	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
37	8	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
38	9	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
39	19	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	21
40	7	8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
41	7	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
42	8	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
43	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
44	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
45	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
Totals	121	93	125	70	69	40	52	45	52	20	35	10	18	—	10	4	7	1	3	1	4	780

1.5. Show that the conditions that the function

$$f(x_1, x_2) = z_0 \exp \{Ax_1^2 + 2Hx_1x_2 + Bx_2^2\}, \qquad -\infty \leqslant x_1, x_2 \leqslant \infty$$

may represent a frequency function are

(a) $A \leqslant 0$
(b) $B \leqslant 0$
(c) $AB - H^2 \geqslant 0$.

Show further that if these conditions are satisfied and the integral of $f(x_1, x_2)$ between $-\infty$ and ∞ for both variates is unity, then

$$z_0 = \frac{\,}{\pi} \begin{matrix} -A & H \\ H & -B \end{matrix}$$

CHAPTER 2

MEASURES OF LOCATION AND DISPERSION

2.1. It has been seen in Chapter 1 that the frequency-distributions occurring in statistical practice vary considerably in general nature. Some are finite in range and some are not. Some are symmetrical and some markedly skew. Some present only a single maximum and others present several. Amid this variety we may, however, discern four general types: (*a*) the symmetrical distribution with a single maximum, such as that of Table 1.7; (*b*) the asymmetrical distribution, or skew distribution, with a single maximum, such as those of Tables 1.8 and 1.9; (*c*) the extremely skew, or J-shaped, distribution, such as that of Table 1.2; and (*d*) the U-shaped distribution, such as that of Table 1.11. To make this classification comprehensive we should have to add a fifth class comprising the miscellaneous distributions not falling into the other four.

The distributions with a single maximum will hereafter be called "unimodal." The synonymous terms "cocked-hat," "single-humped" and one or two others also occur in statistical literature.

2.2. It frequently happens in statistical work that we have to compare two distributions. If one is unimodal and the other J-shaped or multimodal a concise comparison is clearly difficult to make, and in such a case it would probably be necessary to specify both distributions completely. But if both are of the same type (and it is in such cases that comparisons most frequently arise) we may be able to make a satisfactory comparison merely by examining their principal characteristics; e.g. if both are unimodal it might be sufficient to compare (*a*) the whereabouts of some central value, such as the maximum —this, as it were, locates the distributions; (*b*) the degree of scatter about this value —the dispersion; and (*c*) the extent to which the distributions deviate from the symmetrical—the skewness.

The same point emerges when our distributions are specified by some mathematical function. If, for example, we have two distributions of the type

$$dF = y_0 e^{-\frac{(x-m)^2}{2v}}\, dx,$$

symmetrical about $x = m$, a complete comparison can be made by comparing the value of the constants m and v in the distributions. Such constants are called *parameters* of the distribution. This chapter is devoted to a discussion of parameters of location and dispersion.

Measures of Location: the Arithmetic Mean

2.3. There are three groups of measures of location in common use: the means (arithmetic, geometric and harmonic), the median and the mode. We consider them in turn.

The arithmetic mean is perhaps the most generally used statistical measure, and in fact is far older than the science of statistics itself. If the proportional frequency of the values x of a distribution is $f(x)$, the arithmetic mean μ_1' about the point $x = a$ is defined by

$$\begin{aligned}\mu_1' &= \int_{-\infty}^{\infty} (x-a) f(x)\, dx \\ &= \int_{-\infty}^{\infty} (x-a)\, dF \qquad (2.1)\end{aligned}$$

This integral is to be understood in the Stieltjes sense and hence includes summation in the discontinuous case; e.g. the arithmetic mean of a set of discrete values x is their sum divided by the number of values. In formula (2.1) the frequency, in accordance with our usual convention, is expressed as a proportion of the total frequency. If the *actual* frequencies are $g(x)$, totalling N, we have

$$\mu_1' = \frac{1}{N}\int_{-\infty}^{\infty} (x - a)g(x)\, dx$$

in the continuous case, and

$$\mu_1' = \frac{1}{N}\sum_{j=-\infty}^{\infty} (x_j - a)g(x_j)$$

in the discontinuous case. The value of the arithmetic mean thus depends on the value of a, the point from which it is measured. For a mathematically specified distribution the integral (2.1) need not necessarily converge, in which case no arithmetic mean exists.

2.4. The calculation of the arithmetic mean of a numerically specified distribution (i.e. one whose frequency-distribution is given in the form of a numerical table like those of Chapter 1) is a simple process. If there are relatively few values in the population we merely sum them and divide by their total number N. If they are given in the form of a frequency table a more formal procedure is desirable, but the principle is exactly the same. The following example will make the process clear.

Example 2.1

To calculate the arithmetic mean of the population of males distributed according to height of Table 1.7.

Let us note first of all that if b is some other arbitrary variate-value,

$$\begin{aligned}\mu_1' \text{ (about } a) &= \int_{-\infty}^{\infty} (x - a)\, dF \\ &= \int_{-\infty}^{\infty} (x - b)\, dF + \int_{-\infty}^{\infty} (b - a)\, dF \\ &= \mu_1' \text{ (about } b) + b - a \qquad (2.2)\end{aligned}$$

In other words, we can find the mean about any point very simply when we know the mean about any other. In calculating the arithmetic mean we can then take an arbitrary point as origin and transfer to any other desired point afterwards. It is generally convenient to choose this arbitrary point somewhere near the maximum frequency.

One further point arises in grouped data such as this. We do not know *exactly* the variate-values of the individuals within a certain class range. We therefore assume them concentrated at the centre of the interval. Corrections for any distortion thus introduced will be considered in Chapter 3. In fact, no correction is required for the arithmetic mean in the case when the frequency "tails off" at both ends of the distribution.

In the particular case before us we take an arbitrary origin at the centre of the interval 67– inches, i.e. at the point $67\frac{7}{16}$ inches, and measure $\xi (= x - a)$ from that point. Column 2 in Table 2.1 shows the frequency, column 3 the value of ξ and column 4 the value of ξf. We find, having due regard to sign,

$$\Sigma(\xi f) = 8763 - 8584 = 179.$$

Hence the mean about $x = 0$ is $67\frac{7}{16} + \dfrac{179}{8585} = 67{\cdot}46$ inches.

TABLE 2.1

Calculation of the Arithmetic Mean for the Distribution of Table 1.7.

(1) Height, inches.	(2) Frequency f.	(3) Deviation from Arbitrary Value ξ.	(4) Product ξf.
57–	2	− 10	20
58–	4	− 9	36
59–	14	− 8	112
60–	41	− 7	287
61–	83	− 6	498
62–	169	− 5	845
63–	394	− 4	1576
64–	669	− 3	2007
65–	990	− 2	1980
66–	1223	− 1	1223
67–	1329	0	− 8584
68–	1230	+ 1	1230
69–	1063	+ 2	2126
70–	646	+ 3	1938
71–	392	+ 4	1568
72–	202	+ 5	1010
73–	79	+ 6	474
74–	32	+ 7	224
75–	16	+ 8	128
76–	5	+ 9	45
77–	2	+ 10	20
TOTALS	8585	—	+ 8763

Example 2.2

For a distribution specified by a mathematical function, the determination of the mean is a matter of evaluating the integral (2.1), when it exists. For instance, to find the mean of the distribution

$$dF = \frac{1}{B(p,q)}(1-x)^{p-1}\,x^{q-1}\,dx \qquad 0 \leqslant x \leqslant 1$$

we have

$$\mu_1' = \frac{1}{B(p,q)}\int_0^1 (1-x)^{p-1}x^q\,dx$$

$$\frac{B(p,q+1)}{B(p,q)} = \frac{\Gamma(p)\Gamma(q+1)}{\Gamma(p+q+1)}\cdot\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}$$

$$p+q$$

2.5. Apart from its relative simplicity and ease of calculation, qualities which ensure it a firm place in the elementary theory of statistics, the arithmetic mean has a number of properties which make it equally important in advanced theory. For instance :—

(*a*) If in (2.1) we take a equal to μ_1' itself the mean vanishes and consequently the sum over the population of deviations from the arithmetic mean is zero.

(*b*) The mean of a sum is the sum of the means ; i.e. if $f_1, f_2 \ldots f_n$ are the frequency functions of n distributions with means $\mu_1', \nu_1' \ldots \rho_1'$, and if the sum of the frequency functions is g with mean θ', then

$$\begin{aligned}\theta' &= \int_{-\infty}^{\infty} (x-a)\ g(x)\,dx \\ &= \int_{-\infty}^{\infty} (x-a)\{f_1(x)+f_2(x)+\ldots+f_n(x)\}\,dx \\ &= \int_{-\infty}^{\infty} (x-a)f_1(x)\,dx + \int_{-\infty}^{\infty} (x-a)f_2(x)\,dx + \ldots + \int_{-\infty}^{\infty} (x-a)f_n(x)\,dx \\ &= \mu_1' + \nu_1' + \ldots + \rho_1'.\end{aligned}$$

(*c*) We shall see later that mean values are important in the theory of sampling, mainly in virtue of their mathematical tractability, but also because in a certain sense the mean is the best measure of location of some distributions.

The Geometric Mean and the Harmonic Mean

2.6. Two other types of mean are in use in elementary statistics, though they are not of importance in advanced theory.

The geometric mean of N variate-values is the Nth root of their product and is not used if any of the variate-values are negative. For proportional frequencies $f(x)$ we have

$$\left.\begin{aligned} G &= \prod_{j=-\infty}^{\infty} (x_j^{f_j}) \\ \text{or}\qquad \log G &= \sum_{j=-\infty}^{\infty} f_j \log x_j \end{aligned}\right\} \qquad (2.3)$$

and for actual frequencies $g(x)$, totalling N,

$$\left.\begin{aligned} G &= \Pi(x_j^{g_j})^{\frac{1}{N}} \\ \log G &= \frac{1}{N}\Sigma g_j \log x_j \end{aligned}\right\} \qquad (2.4)$$

The harmonic mean of N variate-values is the reciprocal of the arithmetic mean of their reciprocals. In the usual notation

$$\frac{1}{H} = \int_{-\infty}^{\infty} \frac{dF}{x} = \int_{-\infty}^{\infty} \frac{f(x)\,dx}{x}. \qquad (2.5)$$

or, for actual frequencies,

$$\frac{1}{H} = \int_{-\infty}^{\infty} g(x)\,dx \qquad (2.6)$$

Example 2.3

To find the geometric and harmonic means of the distribution

$$dF = \frac{1}{B(p,q)}(1-x)^{p-1}\,x^{q-1}\,dx \qquad 0 \leqslant x \leqslant 1$$

we have

$$\log G = \frac{1}{B(p,q)}\int_0^1 (1-x)^{p-1}\,x^{q-1}\log x\,dx.$$

Now, since by definition

$$\int_0^1 (1-x)^{p-1}\, x^{q-1}\, dx = B(p,\, q)$$

we have, differentiating both sides with respect to q, an operation which is legitimate in virtue of the uniform convergence of the integral and the existence of the resulting expressions,

$$\int_0^1 (1-x)^{p-1}\, x^{q-1} \log x\, dx = \frac{\partial}{\partial q} B(p,\, q).$$

Thus

$$\begin{aligned}\log G &= \frac{1}{B(p,q)}\, \frac{\partial}{\partial q} B(p,\, q) \\ &= \frac{\partial}{\partial q} \log \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \\ &= \frac{\partial}{\partial q} \{\log \Gamma(q) - \log \Gamma(p+q)\}.\end{aligned}$$

The harmonic mean is given by

$$\begin{aligned}\frac{1}{H} &= \frac{1}{B(p,\, q)} \int_0^1 (1-x)^{p-1}\, x^{q-2}\, dx \\ &= \frac{B(p,\, q-1)}{B(p,q)} = \frac{\Gamma(q-1)}{\Gamma(p+q-1)} \cdot \frac{\Gamma(p+q)}{\Gamma(q)} \\ &= \frac{p+q-1}{q-1}\end{aligned}$$

so that

$$H = \frac{q-1}{p+q-1}.$$

We may note that the arithmetic mean, $\frac{}{p+q}$, is greater than the harmonic mean, for

$$\frac{}{p+q} = 1 - \frac{p}{p+q},\quad \frac{q-1}{p+q-1} = 1 - \frac{p}{p+q-1}$$

and therefore

$$\mu_1' > H$$

if

$$\frac{p}{p+q-1} > \frac{p}{p+q}$$

which is clearly so.

2.7. In general it may be shown that for distributions in which the variate-values are not negative

$$H \leqslant G \leqslant \mu_1'. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2.7)$$

Consider in fact the quantity

$$E(t) = \left| \frac{1}{N}(x_1^t + x_2^t + \ .\ .\ .\ x_N^t) \right|^{\frac{1}{t}}$$

where the x's are positive numbers. We shall show that this is an increasing function of t, i.e. $E(t_1) > E(t_2)$ if $t_1 > t_2$. As a trivial case these inequalities may be replaced by equalities, namely if all the x's are equal. We have

$$\begin{aligned}\frac{d}{dt} \log E &= \frac{d}{dt} \log \left(\frac{1}{N} \Sigma x^t\right)^{\frac{1}{t}} \\ &= -\frac{1}{t^2} \log \frac{1}{N} \Sigma x^t + \frac{1}{t}\frac{d}{dt} \log \Sigma x^t.\end{aligned}$$

Hence, for the function

$$F = t^2\frac{d}{dt}\log E$$

we have

$$\frac{dF}{dt} = -\frac{d}{dt}\log \Sigma x^t + \frac{d}{dt}\left\{t\frac{d}{dt}\log \Sigma x^t\right.$$

$$= t\frac{d^2}{dt^2}\log \Sigma x^t$$

$$= \frac{t}{(\Sigma x^t)^2}[\Sigma(x^t \log^2 x)\Sigma(x^t) - \{\Sigma(x^t \log x)\}^2] \tag{2.8}$$

Now in virtue of Schwarz's inequality $\Sigma(a^2)\Sigma(b^2) \geqslant \{\Sigma(ab)\}^2$ the expression in brackets is not negative. Hence $\frac{dF}{dt}$ has the sign of t and F thus has a minimum at $t = 0$. But when $t = 0$, $F = 0$ and thus F must be non-negative. Therefore $\frac{d}{dt}\log E$ is non-negative, and since E is positive $\frac{dE}{dt}$ is non-negative and thus E is a non-decreasing function.

Now in $E(t)$, when $t = 1$ we have the arithmetic mean; when $t = -1$ we have the harmonic mean; and when $t \to 0$ we have the geometric mean, for

$$\lim_{t\to 0} \log E = \lim \frac{\frac{1}{N}\Sigma(x^t)}{t}$$

$$= \lim \frac{1}{N}\Sigma x^t \log$$

$$= \frac{1}{N}\Sigma \log x.$$

Hence the inequality (2.7) follows.

For simplicity we have stated these results for the discontinuous variate. The analysis, however, is easily seen to remain true for Stieltjes integrals and hence is generally valid.

Hereafter when the "mean" is mentioned without qualification, the arithmetic mean is to be understood.

The Median

2.8. The median value is that value of the variate which divides the total frequency into two equal halves, i.e. is the value μ_e such that

$$\int_{-\infty}^{\mu_e} f(x)\,dx = \int_{\mu_e}^{\infty} f(x)\,dx = \tfrac{1}{2}. \tag{2.9}$$

There is some small indeterminacy in this definition when the distribution is discontinuous which may be removed by convention. If there are $(2N + 1)$ members of the population, we take the median to be the value of the $(N+1)$th member. If there are $2N$ we take it to be halfway between the values of the Nth and the $(N + 1)$th. When the distribution is numerically specified in class-intervals there is the usual indeterminacy due to grouping, which may be dealt with in the manner of the following example.

Example 2.4

To find the median value of the distribution of heights considered in Example 2.1.

Half the total frequency of 8585 observations is 4292·5.

There are, up to and including the interval, $66\frac{15}{16}$ inches 3589

leaving 703·5

The frequency in the next interval is 1329

Hence we take the median to be

$$66\tfrac{15}{16} + \frac{703.5}{1329} = 67{\cdot}47 \text{ inches.}$$

The mean (Example 2.1) is 67·46 inches, practically the same.

A graphical method of determining the median is given later in this chapter (**2.13**).

The Mode

2.9. The mode or modal value is that value of the variate exhibited by the greatest number of members of the distribution. If the frequency function is continuous and differentiable it is the solution of

$$f'(x) = \frac{d}{dx}f(x) = 0, \quad f''(x) = \frac{d^2}{dx^2}f(x) < 0 \qquad (2.10)$$

If $f'(x)$ vanishes and $f''(x)$ is greater than zero we have a minimum, and such a point is sometimes called an Antimode.

In numerically specified distributions and discontinuous distributions generally the mode is sometimes difficult to determine exactly. It is essentially a concept related to the continuous frequency function. For example, if the distribution merely consists of an isolated number of values, each of which occurs only once, there is no mode in the sense defined above. Where, however, the number is large enough to permit grouping, there will usually be an interval containing a maximum frequency, and we may regard the mode as lying in that interval. More generally there may be several maxima, in which case the distribution is multimodal. In the height distribution of Table **1**.7, for instance, the mode may be considered as lying somewhere in the interval 67– inches. To estimate its position more accurately it is necessary to fit a continuous curve to the distribution and determine the mode of the curve. The process of fitting will be considered in Chapter 6.

2.10. In a symmetrical distribution the mean, the median and the mode (or in cases such as the U-shaped distribution, the antimode) coincide. For skew distributions they differ. There is an interesting empirical relationship between the three quantities which appears to hold for unimodal curves of moderate asymmetry, namely

$$\text{Mean} - \text{Mode} = 3\ (\text{Mean} - \text{Median}). \qquad (2.11)$$

A mathematical explanation of this relationship has been given by Doodson (1917).

It is a useful mnemonic to observe that the mean, median and mode occur in the same order (or the reverse order) as in the dictionary; and that the median is nearer to the mean than to the mode, just as the corresponding words are nearer together in the dictionary.

In elementary theory the median and the mode have considerable claims to use as measures of location. They are readily interpretable in terms of ordinary ideas—the median is the middle value and the mode is the most popular value—and the median is usually more easily determined than the mean in numerically specified distributions.

What gives the arithmetic mean the greater importance in advanced theory is its superior mathematical tractability and certain sampling properties; but the median has compensating advantages—it is, for instance, less dependent on the scale and the form of the frequency-distribution than the mean—and it seems to deserve more consideration in the advanced theory than it has received.

Quantiles

2.11. The concept of median value can be easily extended to locate the curve more accurately by the use of several parameters. We may, for example, find the three variate-values which divide the total frequency into four equal parts. The middle one of these will be the median itself; the other two are called the lower and upper *quartiles* respectively. Similarly, we may find the nine variate-values which divide the total frequency into ten equal parts—the *deciles*. Generally we may find the $(n - 1)$ variate-values which divide the total frequency into n equal parts—the quantiles. Evidently the knowledge of the quantiles for some fairly high n, such as 10, gives a very good idea of the general form of the frequency-distribution. Even the quartiles and the median are valuable general guides.

2.12. The determination of the quantiles of a numerically specified distribution proceeds as for the median, indeterminacies being resolved by the usual conventions. That of the quantiles of a mathematically specified distribution, say the jth quantile, is a matter of solving the equation

$$\frac{j}{n} = \int_{-c}^{x} dF \tag{2.12}$$

which can be done without difficulty by interpolation when the integral of dF has been tabulated.

Example 2.5

To find the quartiles of the height distribution considered in Example 2.1.

One-quarter of the total frequency is $8585/4 =$	2146·25	
Up to the interval 65– there are	1376	members
leaving	770·25	members
In the next interval there are	990	members
Thus the lower quartile is $64\frac{15}{16} + \dfrac{770{\cdot}25}{990}$	65·71	inches
The upper quartile will be found to be	69·21	inches
We have already found (Example 2.4) that the median is	67·47	inches

Denoting the quartiles by Q_1 and Q_3 we see that

$$Q_1 - \mu_e = -\ 1{\cdot}76 \text{ inches}$$
$$Q_3 - \mu_e = 1{\cdot}74 \text{ inches}$$

so that the median is almost half-way between the quartiles, an indication of the symmetry of the distribution.

The Distribution Curve or Ogive of Galton

2.13. The quantiles may also be determined graphically. Suppose we plot x, the variate, along a horizontal axis and $\Sigma f(x)$, the cumulated frequency up to and including

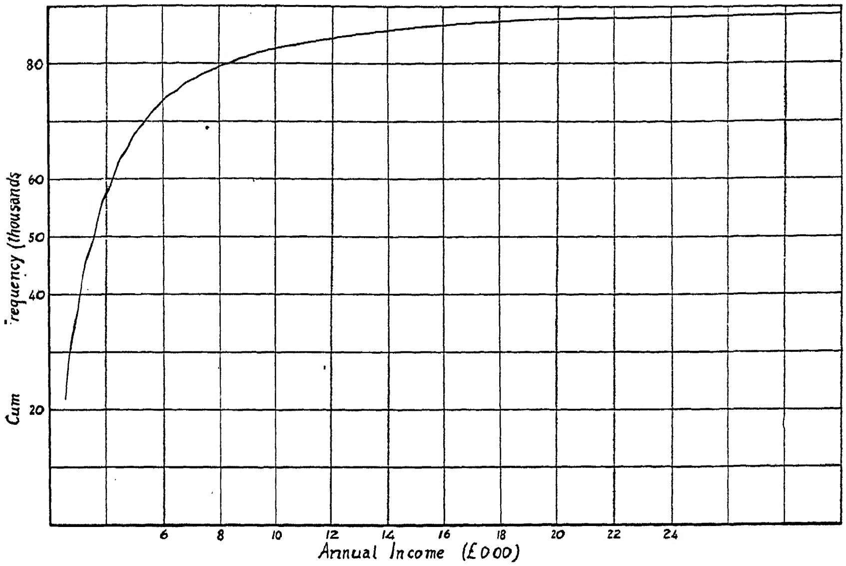


FIG. 2.1. Distribution Curve of the Data of Table 1.2.

x, along the perpendicular y-axis. We then get a series of points through which, in general, a smooth curve may be drawn. This curve, as is evident from its definition, is

$$y = F(x),$$

i.e. the graph of the distribution function. It is sometimes called the graduation curve, or Galton's ogive (though it is only shaped like an ogive in certain cases such as that of a 'unimodal symmetrical curve). We shall use the expression "distribution curve."

Fig. 2.1 illustrates the distribution curve for the J-shaped distribution of Table 1.2, and Fig. 2.2 that for the unimodal symmetrical distribution of Table 1.7. A freehand curve has been drawn in both cases.

Curves of this kind can be used to determine the quantiles. In fact, to find the median, we merely have to find the abscissa corresponding to the ordinate $N/2$,

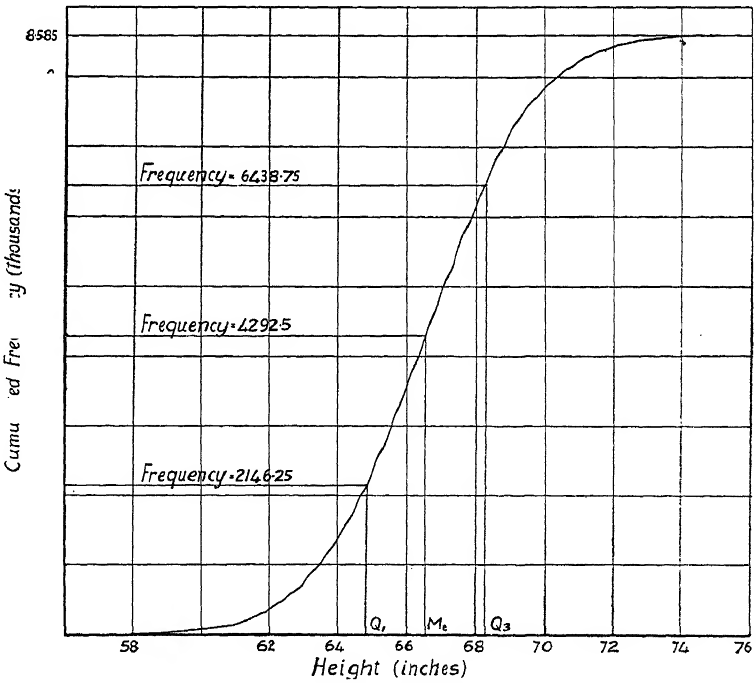


FIG. 2.2. Distribution Curve of the Data of Table 1.7.
(Heights shown to correspond to entries in the Table, e.g. cumulated frequency at 64 inches is the frequency up to and including the range 64– and therefore up to $64\frac{15}{16}$ inches.)

and so on. The positions of the quartiles and the median are shown in Fig. 2.2, and the reader may care to compare the values obtained by reading the graph by eye with those given in Example 2.5.

Measures of Dispersion

2.14. We now proceed to consider the quantities which have been proposed to measure the dispersion of a distribution. They fall into three groups:—

(*a*) Measures of the distance (in terms of the variate) between certain representative values, such as the range, the interdecile range or the interquartile range.

(*b*) Measures compiled from the deviations of every member of the population from some central value, such as the mean deviation from the mean, the mean deviation from the median, and the standard deviation.

(*c*) Measures compiled from the deviations of all the members of the population among themselves, such as the mean difference.

In advanced theory the outstandingly important measure is the standard deviation; but they all require some mention.

Range and Interquantile Differences

2.15. The range of a distribution is the difference of the greatest and least variate-values borne by its members. As a descriptive parameter of a population it has very little use. A knowledge of the whereabouts of the end values obviously tells little about the way the bulk of the distribution is condensed inside the range; and for distributions of infinite range it is obviously wholly inappropriate.

More useful rough-and-ready measures may be obtained from the quantiles, and there are two such in general use. The interquartile range is the distance between the upper and lower quartiles, and is thus a range which contains one-half the total frequency. The interdecile range (or perhaps, more accurately the 1–9th interdecile range) is the distance between the first and the ninth decile. Both these measures evidently give some approximate idea of the "spread" of a distribution, and are easily calculable. For this reason they are fairly generally used in elementary descriptive statistics. In advanced theory they suffer from the disadvantage of being difficult to handle mathematically in the theory of sampling.

Mean Deviations

2.16. The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean. We have seen (2.5) that the sum of these deviations taken with appropriate sign is zero. We may however write

$$\delta_1 = \int_{-\infty}^{\infty} | x - \mu_1 | \, dF \tag{2.13}$$

where the deviations are now taken absolutely, and define δ_1 to be a coefficient of dispersion. We shall call it the mean deviation about the mean.

Similarly for the median μ_e we may write

$$= \int_{-c}^{\infty} x - \mu_e | \, dF \tag{2.14}$$

and call δ_2 the mean deviation about the median.

In future the words "mean deviation" alone will be taken to refer to the mean deviation about the mean.

Both these measures have merits in elementary work, being fairly easily calculable. Once again, however, they are practically excluded from advanced work by their intractability in the theory of sampling.

Standard Deviation

2.17. We have seen that the mean about an arbitrary point a is given by

$$\mu_1' = \int_{-\infty}^{\infty} (x - a)\, dF.$$

We may, by analogy with the terminology of Statics, call this the first moment, and define the second moment by

$$\mu_2' = \int_{-\infty}^{\infty} (x - a)^2\, dF \quad . \quad . \quad . \quad . \quad . \quad (2.15)$$

The second moment about the mean is written without the prime, thus:

$$\mu_2 = \int_{-\infty}^{\infty} (x - \mu_1')^2\, dF \quad . \quad . \quad . \quad . \quad . \quad (2.16)$$

and is called the Variance. The positive square root of the variance is called the standard deviation, and usually denoted by σ, so that we have

$$\sigma = + \sqrt{\mu_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.17)$$

The variance is thus the mean of the *squares* of deviations from the mean. The device of squaring and then taking the square root of the resultant sum in order to obtain the standard deviation may appear a little artificial, but it makes the mathematics of the sampling theory very much simpler than is the case, for example, with the mean deviation.

The calculation of the variance and the standard deviation proceeds by an easy extension of the methods used for the mean. In particular, if b is some arbitrary value

$$\begin{aligned} \mu_2' \text{ (about } a) &= \int_{-\infty}^{\infty} (x - a)^2\, dF \\ &= \int_{-\infty}^{\infty} \{(x - b)^2 + 2(b - a)(x - b) + (b - a)^2\} dF \\ &= \mu_2' \text{ (about } b) + 2(b - a)\mu_1' \text{ (about } b) + (b - a)^2 \end{aligned} \quad (2.18)$$

If now b is the mean we have

$$\mu_2' = \mu_2 + (\mu_1' - a)^2$$

or

$$\mu_2 = \mu_2' - (\mu_1' - a)^2 \quad . \quad . \quad . \quad . \quad . \quad (2.19)$$

Thus the variance can easily be found from the second moment about an arbitrary point, which can be selected to simplify the calculations.

Example 2.6

To find the mean deviation and the standard deviation for the distribution of men according to height considered in Example 2.1 (Table 1.7).

In the case of the mean deviation for a grouped distribution, the sum of deviations should first be calculated from the centre of the class-interval in which the mean lies and then reduced to the mean as origin. It so happens that in Table 2.1 the mean fell in the interval taken as origin, so that the preliminary arithmetic already exists in the Table.

The sum of positive deviations is 8763 and that of negative deviations $-$ 8584. Hence the sum of deviations regardless of sign is 17.347, the unit being the class-interval and the origin the centre of the interval.

To reduce to the mean as origin, we note that if the number of observations below the mean is N_1 and the number above the mean is N_2, and $d = \mu_1' - a$, we have to add N_1d to the sum of deviations about the centre of the interval and subtract N_2d. In this case $d = 0{\cdot}02$ (Example 2.1), $N_1 = 4918$, $N_2 = 3667$. Hence we add $(4918 - 3667)0{\cdot}02 = 25$. Hence the mean deviation

$$\delta_1 = \frac{17,347 + 25}{8585} = 2{\cdot}02 \text{ inches.}$$

For the standard deviation some further calculation is required, as shown in Table 2.2.

TABLE 2.2

Calculation of the Standard Deviation for the Distribution of Table 1.7.
(Some preliminary calculation already carried out in Table 2.1.)

(1) Height, inches.	(2) Frequency f.	(3) Deviation ξ.	(4) $\xi^2 f$.
57–	2	– 10	200
58–	4	– 9	324
59–	14	– 8	896
60–	41	– 7	2,009
61–	83	– 6	2,988
62–	169	– 5	4,225
63–	394	– 4	6,304
64–	669	– 3	6,021
65–	990	– 2	3,960
66–	1223	– 1	1,223
67–	1329	0	0
68–	1230	1	1,230
69–	1063	2	4,252
70–	646	3	5,814
71–	392	4	6,272
72–	202	5	5,050
73–	79	6	2,844
74–	32	7	1,568
75–	16	8	1,024
76–	5	9	405
77–	2	10	200
TOTALS	8585	—	56,809

Column (4) shows the sum $\Sigma\xi^2 f$, where f is the actual frequency. We then have, for the second moment about the arbitrary origin

$$\mu_2 = \frac{56,809}{8585} \quad 6{\cdot}6172.$$

We have already found in Example 2.1 that

$$\mu_1 \cdot a = \frac{179}{8585} = 0{\cdot}0209.$$

Hence, in virtue of (2.19)

$$\mu_2 = 6{\cdot}6172 - (0{\cdot}0209)^2$$
$$= 6{\cdot}6168$$
$$\sigma = \sqrt{\mu_2} = 2{\cdot}57 \text{ inches.}$$

It may be noted that the mean deviation is about 80 per cent. of the standard deviation. This relationship often holds approximately for unimodal curves approaching symmetry. The reason will become apparent when we study the so-called "normal" distribution in Chapter 5.

Example 2.7

To find the variance of the distribution

$$dF = \frac{1}{B(p, q)}(1 - x)^{p-1} x^{q-1}\, dx, \qquad 0 \leqslant x \leqslant 1.$$

We have, about the origin,

$$\mu_2' = \frac{1}{B(p, q)}\int_0^1 (1 - x)^{p-1} x^{q+1}\, dx$$
$$= \frac{B(p, q + 2)}{B(p, q)} = \frac{(q + 1)q}{(p + q + 1)(p + q)}.$$

We have already found (Example 2.2) that

$$\mu_1' = \frac{q}{p + q}.$$

Thus

$$\mu_2 = \mu_2' - \mu_1'^2$$
$$= \frac{(q + 1)q}{(p + q + 1)(p + q)} - \frac{q^2}{(p + q)^2}$$
$$= \frac{pq}{(p + q + 1)(p + q)^2}.$$

Sheppard's Corrections

2.18. The treatment of the values of a grouped frequency-distribution as if they were concentrated at the mid-points of intervals is an approximation, and in certain circumstances it is possible to make corrections for any distortion introduced thereby. These so-called "Sheppard's corrections" will be discussed at length in the next chapter, but at this stage we may indicate without proof the appropriate correction for the second moment.

If the distribution is continuous and has high order contact with the variate-axis at its extremities, i.e. if it "tails off" slowly, the crude second moment calculated from grouped frequencies should be corrected by subtracting from it $h^2/12$, where h is the width of the interval. For example, in the height data of Example 2.6, we have $h = 1$, and the corrected second moment is

$$6{\cdot}6168 - 0{\cdot}0833 = 6{\cdot}5335.$$

The corrected value of σ is $\sqrt{6{\cdot}5335} = 2{\cdot}56$, as against an uncorrected value of $2{\cdot}57$.

Mean Difference

2.19. The coefficient of mean difference (not to be confused with mean deviation) is defined by

$$\Delta_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} | x - y | \, dF(x) \, dF(y)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} | x - y | f(x) f(y) \, dx \, dy \quad . \quad . \quad . \quad . \quad . \quad (2.20)$$

In the discontinuous case two different formulae arise. We have either

$$\Delta_1' = \frac{1}{N(N-1)} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} | x_j - x_k | f(x_j) f(x_k), \quad j \neq k \qquad (2.21)$$

the mean difference without repetition, or

$$\Delta_1 = \frac{1}{N^2} \sum_{j=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} | x_j - x_k | f(x_j) f(x_k), \quad . \quad . \quad (2.22)$$

the mean difference with repetition. The difference lies only in the divisor and is unimportant if N is large.

The mean difference is the average of the differences of all the possible pairs of variate-values, taken regardless of sign. In the coefficient with repetition each value is taken with itself, adding of course nothing to the sum of deviations, but resulting in the total number of pairs being N^2. In the coefficient without repetition only different values are taken, so that the number of pairs is $N(N - 1)$. Hence the divisors in (2.21) and (2.22).

2.20. The mean difference, which is due to Gini (1912), has a certain theoretical attraction, being dependent on the spread of the variate-values among themselves and not on the deviations from some central value. It is, however, more difficult to compute than the standard deviation, and the appearance of the absolute values in the defining equations indicates, as for the mean deviation, the appearance of difficulties in the theory of sampling. It might be thought that this inconvenience could be overcome by the definition of a coefficient

$$E^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - y)^2 \, dF(x) \, dF(y).$$

This, however, is nothing but twice the variance.

For

$$\begin{aligned} E^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dF(x) \, dF(y) \; \{x^2 - 2xy + y^2\} \\ &= \int_{-\infty}^{\infty} x^2 \, dF(x) \int_{-\infty}^{\infty} dF(y) - 2\int_{-\infty}^{\infty} x \, dF(x) \int_{-\infty}^{\infty} y \, dF(y) \\ &\qquad + \int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{\infty} y^2 \, dF(y) \\ &= 2\mu_2' - 2\mu_1'^2 \\ &= 2\mu_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.23) \end{aligned}$$

This interesting relation shows that the variance may in fact be defined as half the mean square of all possible variate differences, that is to say, without reference to deviations from a central value, the mean.

Coefficients of Variation: Standard Measure

2.21. The foregoing measures of dispersion have all been expressed in terms of units of the variate. It is thus difficult to compare dispersions in different populations unless the units happen to be identical; and this has led to a search for measures which shall be independent of the variate scale, that is to say, shall be pure numbers.

Several coefficients of this kind may be constructed, such as the $\frac{\text{mean deviation}}{\text{mean}}$ or $\frac{\text{mean deviation}}{\text{median}}$. Only two have been used at all extensively in practice, Karl Pearson's coefficient of variation, defined by

$$v = 100\frac{\sigma}{\mu_1'} \qquad (2.24)$$

and Gini's coefficient of concentration, defined by

$$G = \frac{\Delta_1}{2\mu_1'}. \qquad (2.25)$$

Both these coefficients suffer from the disadvantage of being affected very much by μ_1', the value of the mean measured from some arbitrary origin, and are hardly suitable for advanced work.

2.22. For our purposes, comparability may be attained in a somewhat different way. Let us take σ itself as a new unit and express the frequency function in terms of a new variable ξ related to x by

$$\xi = \frac{x - \mu_1'}{\sigma} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2.26)$$

Any distribution expressed in this way has zero mean and unit variance. It is then said to be expressed in standard measure. Two distributions in standard measure can be readily compared in regard to form, skewness, and other qualities, though not of course in regard to mean and variance.

Concentration

2.23. Gini's coefficient of concentration arises in a natural way from the following approach :—

Writing, as usual

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \qquad (2.27)$$

let us define

$$\Phi(x) = \frac{1}{\mu_1'}\int_{-\infty}^{x} x\,f(x)\,dx \quad . \quad . \quad . \quad . \quad . \qquad (2.28)$$

$\Phi(x)$ exists, of course, only if μ_1' exists. Just as $F(x)$ varies from 0 to 1, $\Phi(x)$ varies from 0 to 1 provided that the origin is taken to the left of the start of the frequency-distribution, which we shall assume to be so. $\Phi(x)$ may be called the *incomplete* first moment.

Now (2.27) and (2.28) may be regarded as defining a relationship between the variables F and Φ in terms of parametric equations in x.* The curve whose ordinate

* The definition of curves by parametric equations will be found treated in most textbooks of differential calculus. The term "parameter" in this connection is usual in mathematics, but is not to be confused with the more special statistical parameter as defined in **2.2**.

and abscissa are Φ and F is called the curve of concentration. Such a curve is shown in Fig. 2.3.

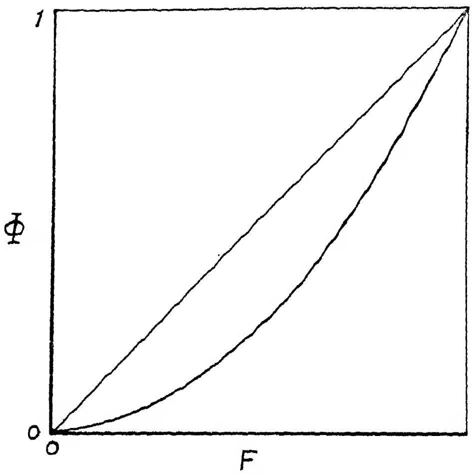


FIG. 2.3.—Curve of Concentration.

The curve of concentration must be convex to the F-axis, for we have

$$\mu_1' \frac{d\Phi}{dF} = \frac{xf(x)}{f(x)} = x,$$

which is positive since our origin is taken to the left of the start of the distribution.

$$\mu_1' \frac{d^2\Phi}{dF^2} = \frac{dx}{dF} = \frac{1}{f(x)} = \text{positive}.$$

Thus the tangent to the curve makes a positive acute angle with the F-axis, and the angle increases as F increases; in other words, the curve is convex to the F-axis.

The area between the concentration curve and the line $F = \Phi$ is called the area of concentration. We proceed to show that it is equal to one-half the coefficient of concentration.

In fact, we have from Fig. 2.3

$$2 \text{ (area of concentration)} = \int_0^1 F\, d\Phi - \int_0^1 \Phi\, dF$$

and thus

$$\begin{aligned} 2\mu_1' \text{ (area)} &= \int_{-\infty}^{\infty} F(x)x\, dF(x) - \mu_1' \int_{-\infty}^{\infty} \Phi(x)\, dF(x) \\ &= \int_{-\infty}^{\infty} x\, dF(x) \int_{-\infty}^{x} dF(y) - \int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{x} y\, dF(y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{x} (x-y)\, dF(x)\, dF(y). \end{aligned}$$

Now $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-y)\, dF(x)\, dF(y) = 0$, and hence

$$\begin{aligned} 2\mu_1' \text{ (area)} &= \tfrac{1}{2}\left[\int_{-\infty}^{\infty} \int_{-\infty}^{x} (x-y)\, dF(x)\, dF(y) + \int_{-\infty}^{\infty} \int_{x}^{\infty} (y-x)\, dF(x)\, dF(y)\right] \\ &= \tfrac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\, x - y \,|\ dF(x)\, dF(y) \end{aligned}$$

A

2

Thus

$$\text{area of concentration} = \frac{1}{2}\frac{\Delta_1}{2\mu_1'} = \frac{1}{2}G, \text{ the coefficient of concentration.}$$

2.24. Various methods have been given for calculating the mean difference. The following is probably the simplest, particularly for distributions specified in equal group-intervals.

Let us, without loss of generality, take an origin at the start of the distribution. We may then write

$$\sum_{j=1}^{N} \sum_{k=1}^{N} | x_j - x_k | = 2\Sigma'(x_j - x_k)$$

the summation Σ' being taken over values such that $j \geqslant k$. We have also

$$x_j - x_k = (x_j - x_{j-1}) + (x_{j-1} - x_{j-2}) + \ldots + (x_{k+1} - x_k).$$

Thus

$$\Sigma'(x_j - x_k) = \sum_{h=1}^{N-1} C_h(x_{h+1} - x_h)$$

where C_h is the number of terms of type $(x_j - x_k)$ in Σ' containing $x_{h+1} - x_h$. Since h is the number of values of j less than or equal to h (the origin being at the start of the distribution) and $N - h$ the number greater than or equal to $h + 1$, we have $C_h = h(N - h)$, and thus

$$\Delta_1 = \frac{2}{N^2}\Sigma'(x_j - x_k)$$

$$= \frac{2}{N^2}\sum_{h=1}^{N-1} h(N - h)(x_{h+1} - x_h). \quad . \quad . \quad (2.29)$$

This form is particularly useful if all the intervals are equal. F_h being the distribution function of x_h we then have

$$\Delta_1 = \frac{2}{N^2}\sum_{h=1}^{N-1} (NF_h)(N - NF_h)$$

$$= 2\sum_{h=1}^{N-1} F_h(1 - F_h) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.30)$$

If the actual cumulated frequency for x_h is G_h we have

$$\Delta_1 = \frac{2}{N^2}\sum_{h=1}^{N-1} G_h(N - G_h) \quad . \quad (2.31)$$

the most convenient form in practice.

Example 2.8

Returning once more to the height distribution considered in previous examples, we may calculate $\Sigma G_h(N - G_h)$ as in the Table overleaf.

TABLE 2.3

Calculation of the Mean Difference for the Height Distribution of Table 1.7.

Height, inches.	Frequency.	G_h.	$N - G_h$.	$G_h(N - G_h)$.
57–	2	2	8583	17,166
58–	4	6	8579	51,474
59–	14	20	8565	171,300
60–	41	61	8524	519,964
61–	83	144	8441	1,215,504
62–	169	313	8272	2,589,136
63–	394	707	7878	5,569,746
64–	669	1376	7209	9,919,584
65–	990	2366	6219	14,714,154
66–	1223	3589	4996	17,930,644
67–	1329	4918	3667	18,034,306
68–	1230	6148	2437	14,982,676
69–	1063	7211	1374	9,907,914
70–	646	7857	728	5,719,896
71–	392	8249	336	2,771,664
72–	202	8451	134	1,132,434
73–	79	8530	55	469,150
74–	32	8562	23	196,926
75–	16	8578	7	60,046
76–	5	8583	2	17,166
77–	2	8585	—	—
TOTALS	8585			105,990,850

We have, from (2.31), for the mean difference with repetition,

$$\Delta_1 = \frac{2 \times 105,990,850}{8585^2}$$

$$= 2{\cdot}88 \text{ inches}$$

as against a mean deviation of 2·02 inches and a standard deviation of 2·57 inches (Example 2.6). There is, of course, nothing inconsistent in the difference between these values. The coefficients are different in nature, and there is no reason why their numerical values in any particular case should approach equality.

NOTES AND REFERENCES

The relationship between mean, median and mode expressed in equation (2.11) was discussed from the mathematical point of view by Doodson (1917), who showed that it holds as a first approximation for continuous distributions deviating only moderately from symmetry.

It was shown by Dunham Jackson (1921) that the indeterminacy in the definition of the median can be removed by a more sophisticated mathematical approach. He showed that for N values $x_1 \ldots x_N$ the sum $\sum_{j=1}^{N} | \xi - x_j |^p$, considered as a function of ξ, has

a minimum for some unique ξ_p if $p > 1$; and further that as $p \rightarrow 1$, ξ_p tends to some unique value, which may be defined as the median.

The proof of the increasing character of the function $F(t)$ of **2.7** is due to Norris (1935), who gives references to earlier proofs.

The work of the Italian school on concentration does not appear to have been treated in English books. The fundamental memoir is that of Gini (1912), who has returned to the subject in subsequent papers, many of them in *Metron*. For methods of calculating the mean difference, see de Finetti and Paciello (1930).

de Finetti, B., and Paciello, U. (1930), "Sui metodi proposti per il calcolo della differenza media," *Metron*, **8**, part 3, 89.

Doodson, A. T. (1917), "Relation of Mode, Median and Mean in Frequency Curves," *Biometrika*, **11**, 425.

Gini, C. (1912), "Variabilità e Mutabilità," *Studi Economico-Giuridici della R. Università di Cagliari*, Anno 3, part 2, p. 80.

Gini, C., and Galvani, L. (1929), "Di taluni estensioni dei concetti di media ai caratteri qualitativi," *Metron*, **8**, parts 1–2, 3.

Jackson, Dunham (1921), "Note on the median of a set of numbers," *Bull. Amer. Math. Soc.*, **27**, 160.

Norris, Nilan (1935), "Inequalities among averages," *Ann. Math. Stats.*, **6**, 27.

EXERCISES

2.1. Show that the mean deviation about an arbitrary point is least when that point is the median.

2.2. Show that the mean (about the origin) of the discontinuous distribution whose frequencies at 0, 1, 2, . . . r, . . . are

$$^{n}, \ e^{-m}\frac{m}{1!}, \ e^{-m}\frac{m^2}{2!}, \ . \quad . \ e^{-m}\frac{m^r}{r!}, \ . \ . \ .$$

is m, and that the variance is also m.

2.3. Show that, if deviations are small compared with the value of the mean, we have approximately, for the Geometric and Harmonic means,

$$G = \mu_1'\left(1 - \frac{1}{2}\frac{\sigma^2}{\mu_1'^2}\right)$$

$$H = \mu_1'\left(1 - \frac{\sigma^2}{\mu_1'^2}\right)$$

and hence that

$$\mu_1' - 2G + H = 0.$$

2.4. Show that the mean deviation about the mean is not greater than the standard deviation.

2.5. Show that for the "rectangular" population

$$dF = dx, \qquad 0 \leqslant x \leqslant 1$$

$$\mu_1' \text{ (about the origin)} = \tfrac{1}{2}$$

$$\mu_2 = \tfrac{1}{12}$$

$$\text{mean deviation} = \tfrac{1}{4}$$

$$\Delta_1 = \tfrac{1}{3}.$$

2.6. Show that for the distribution

$$dF = y_0 e^{-\frac{x}{\sigma}}\, dx, \qquad 0 \leqslant x \leqslant \infty$$

the mean, standard deviation and mean difference are all equal to σ; and that the inter-quartile range is $\sigma \log_e 3$.

2.7. Show that for the distribution

$$dF = y_0 e^{-\frac{\chi^2}{2}} \chi^{\nu-1}\, d\chi, \qquad 0 \leqslant \chi \leqslant \infty$$

$$\mu_1' \text{ (about the origin)} = \Gamma\left(\frac{\nu+1}{2}\right) \Big/ \Gamma\left(\frac{\nu}{2}\right)$$

$$\mu_2 = \frac{\nu}{2}.$$

2.8. Show that if a range of six times the standard deviation contains at least 18 class-intervals, Sheppard's correction will make a difference of less than 0·5 per cent. in the uncorrected value of the standard deviation.

2.9. Show that for a continuous distribution

$$\Delta_1 = 2\int_{-\infty}^{\infty} F(x)\{1 - F(x)\}\, dx.$$

2.10. If the variate-values of a distribution are $x_1 \ldots x_N$ in ascending order of magnitude and

$$s_r = \sum_{j=1}^{r} x_j \qquad U = \sum_{r=1}^{N} s_r$$

$$t_r = \sum_{j=1}^{r} x_{N-j+1} \qquad V = \sum_{r=1}^{N} t_r$$

then

$$\Delta_1 = \frac{2}{N^2}(V - U)$$

$$= \frac{2}{N^2}\{N(N+1)\mu_1' - 2U\}.$$

CHAPTER 3

MOMENTS AND CUMULANTS

Definition of Moments

3.1. In the previous chapter we defined the first moment (arithmetic mean) about an arbitrary point a by the Stieltjes integral

$$\mu'_1 = \int_{-\infty}^{\infty} (x - a)dF \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

and the second moment about the point by

$$\mu'_2 = \int_{-\infty}^{\infty} (x - a)^2 dF. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.2)$$

In generalisation of these equations we may define a series of coefficients μ'_r, $r = 1, 2 \ldots$, by the relation

$$\mu'_r = \int_{-\infty}^{\infty} (x - a)^r dF. \quad . \quad . \quad (3.3)$$

μ'_r is called the moment of order r about the point a. When a is the mean μ'_1 we write the moment without the prime,

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu'_1)^r dF. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.4)$$

In particular

$$\mu_1 = 0,$$

and we may also define a moment of zero order

$$\mu'_0 = \mu_0 = \int_{-\infty}^{\infty} dF = 1.$$

It is assumed that when reference is made to the rth moment of a particular distribution, the appropriate integral (3.3) converges for that distribution. As will be seen later, some of the theoretical distributions encountered in statistics do not possess moments of all orders; some, in fact, possess only a few moments of low order, and one or two do not possess any, except of course the moment of order zero.

3.2. If a and b are two variate-values, let $b - a = c$ and denote the moments about a and b by $\mu'(a)$ and $\mu'(b)$ respectively. Then we have, by the binomial theorem,

$$\begin{aligned}(x - a)^r &= (x - b + b - a)^r = (x - b + c)^r \\ &= \sum_{j=0}^{r} \binom{r}{j} (x - b)^{r-j} c^j.\end{aligned}$$

Hence

$$\begin{aligned}\mu'_r(a) &= \int_{-\infty}^{\infty} (x - a)^r dF \\ &= \int_{-\infty}^{\infty} \sum_{j=0}^{r} \binom{r}{j} (x - b)^{r-j} c^j dF \\ &= \sum_{j=0}^{r} \binom{r}{j} c^j \int_{-\infty}^{\infty} (x - b)^{r-j} dF \\ &= \sum_{j=0}^{r} \binom{r}{j} \mu'_{r-j}(b)\, c^j \quad . \qquad (3.5)\end{aligned}$$

This equation gives the rth moment about a in terms of the rth and lower moments about b. It may be written in a symbolic form which will be found to provide a useful mnemonic, namely

$$\mu'_r(a) = \{\mu'(b) + c\}^r$$

with the convention that the expression on the right is to be expanded binomially and the form $\{\mu'(b)\}^j$ replaced by $\mu'_j(b)$.

The equation (3.5) is of particular importance if one of the values a or b is the mean of the distribution. In this case we have

$$\mu'_r = \sum_{j=0}^{r} \binom{r}{j} \mu_{r-j}\, \mu_1'^{j} \quad . \qquad (3.6)$$

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} \mu'_{r-j}\, (-\mu'_1)^j \quad . \qquad (3.7)$$

In particular

$$\left.\begin{aligned} \mu'_2 &= \mu_2 + \mu_1'^2 \\ \mu'_3 &= \mu_3 + 3\mu'_1\mu_2 + \mu_1'^3 \\ \mu'_4 &= \mu_4 + 4\mu'_1\mu_3 + 6\mu_1'^2\mu_2 + \mu_1'^4 \end{aligned}\right\} \quad . \qquad (3.8)$$

and

$$\left.\begin{aligned} \mu_2 &= \mu'_2 - \mu_1'^2 \\ \mu_3 &= \mu'_3 - 3\mu'_1\mu'_2 + 2\mu_1'^3 \\ \mu_4 &= \mu'_4 - 4\mu'_1\mu'_3 + 6\mu_1'^2\mu'_2 - 3\mu_1'^4 \end{aligned}\right\} \quad . \qquad (3.9)$$

Calculation of Moments

3.3. For a distribution specified numerically in a frequency table the calculation of moments of third and higher orders is akin to that of the first and second moments. For grouped data (high order moments are hardly ever required for ungrouped data) the observations are regarded as concentrated at the mid-points of intervals; a convenient arbitrary origin a is chosen, the moments about a calculated, and then if necessary the moments about the mean are ascertained from (3.6) or (3.7). The effect of grouping may be corrected for in certain cases.

In practice numerical moments of order higher than the fourth are rarely required, being so sensitive to sampling fluctuations that values computed from moderate numbers of observations are subject to a large margin of error.

There are two methods in general use for arriving at the moments about an arbitrary origin. The first is an immediate generalisation of the methods used in Chapter 2 for the first two moments. The second will be considered in **3.10** in connection with factorial moments.

Example 3.1

To find the first four moments about the mean of the distribution of Australian marriages of Table 1.8.

Until the last stage we work in units of three years, the variate interval. A working mean is taken at 28·5 years. To check the arithmetic we use an identity of type

$$(x + 1)^3 = x^3 + 3x^2 + 3x + 1$$
$$(x + 1)^4 = x^4 + 4x^3 + 6x^2 + 4x + 1.$$

Thus, for instance, the value of $g(x)(x+1)^r$ is found in addition to that of $g(x)x^r$ and the two checked by identities such as

$$\Sigma g(x)(x+1)^3 = \Sigma g(x)x^3 + 3\Sigma g(x)x^2 + 3\Sigma g(x)x + \Sigma g(x),$$

$g(x)$ being the actual frequencies. The arithmetic work is shown in Table 3.1.

TABLE 3.1

Calculation of the First Four Moments of the Distribution of Marriages of Table 1.8.

Mid-value of Intervals, Years.	g.	x.	xg.	$(x+1)g$.	x^2g.	$(x+1)^2g$.	x^3g.	$(x+1)^3g$.	x^4g.	$(x+1)^4g$.
16·5	294	−4	− 1,176	− 882	4,704	2,646	− 18,816	− 7,938	75,264	23,814
19·5	10,995	−3	− 32,985	−21,990	98,955	43,980	−296,865	− 87,960	890,595	175,920
22·5	61,001	−2	−122,002	−61,001	244,004	61,001	−488,008	− 61,001	976,016	61,001
25·5	73,054	−1	− 73,054	−83,873	73,054	—	− 73,054	−156,899	73,054	—
28·5	56,501	0	−229,217	56,501	—	56,501	−876,743	56,501	—	56,501
31·5	33,478	1	33,478	66,956	33,478	133,912	33,478	267,824	33,478	535,648
34·5	20,569	2	41,138	61,707	82,276	185,121	164,552	555,363	329,104	1,666,089
37·5	14,281	3	42,843	57,124	128,529	228,496	385,587	913,984	1,156,761	3,655,936
40·5	9,320	4	37,280	46,600	149,120	233,000	596,480	1,165,000	2,385,920	5,825,000
43·5	6,236	5	31,180	37,416	155,900	224,496	779,500	1,346,976	3,897,500	8,081,856
46·5	4,770	6	28,620	33,390	171,720	233,730	1,030,320	1,636,110	6,181,920	11,452,770
49·5	3,620	7	25,340	28,960	177,380	231,680	1,241,660	1,853,440	8,691,620	14,827,520
52·5	2,190	8	17,520	19,710	140,160	177,390	1,121,280	1,596,510	8,970,240	14,368,590
55·5	1,655	9	14,895	16,550	134,055	165,500	1,206,495	1,655,000	10,858,455	16,550,000
58·5	1,100	10	11,000	12,100	110,000	133,100	1,100,000	1,464,100	11,000,000	16,105,100
61·5	810	11	8,910	9,720	98,010	116,640	1,078,110	1,399,680	11,859,210	16,796,160
64·5	649	12	7,788	8,437	93,456	109,681	1,121,472	1,425,853	13,457,664	18,536,089
67·5	487	13	6,331	6,818	82,303	95,452	1,069,939	1,336,328	13,909,207	18,708,592
70·5	326	14	4,564	4,890	63,896	73,350	894,544	1,100,250	12,523,616	16,503,750
73·5	211	15	3,165	3,376	47,475	54,016	712,125	864,256	10,681,875	13,828,096
76·5	119	16	1,904	2,023	30,464	34,391	487,424	584,647	7,798,784	9,938,999
79·5	73	17	1,241	1,314	21,097	23,652	358,649	425,736	6,097,033	7,663,248
82·5	27	18	486	513	8,748	9,747	157,464	185,193	2,834,352	3,518,667
85·5	14	19	266	280	5,054	5,600	96,026	112,000	1,824,494	2,240,000
88·5	5	20	100	105	2,000	2,205	40,000	46,305	800,000	972,405
TOTALS	301,785	—	318,049	474,490	2,155,838	2,635,287	13,675,105	19,991,056	137,306,162	202,091,751

From this table we find

$$\begin{aligned}\Sigma(xg) &= 88,832\\ \Sigma(x^2g) &= 2,155,838\\ \Sigma(x^3g) &= 12,798,362\\ \Sigma(x^4g) &= 137,306,162.\end{aligned}$$

The values will be found to check and we have, about the working mean, on dividing by the total frequency 301,785,

$$\begin{aligned}\mu_1' &= 0{\cdot}294,355,253\\ \mu_2' &= 7{\cdot}143,622,115\\ \mu_3' &= 42{\cdot}408,873,867\\ \mu_4' &= 454{\cdot}980,075,219.\end{aligned}$$

For the moments about the mean, substitution in equations (3.9) gives

$$\mu_2 = \quad 7{\cdot}056{,}977$$
$$\mu_3 = \quad 36{\cdot}151{,}595$$
$$\mu_4 = 408{\cdot}738{,}210.$$

These are expressed in class-intervals, which are units of three years. To express the results in units of one year we multiply the rth moment by 3^r, e.g.

$$\mu_2 = 7{\cdot}056{,}977 \times 9 = 63{\cdot}512{,}79.$$

3.4. If a distribution is specified mathematically the determination of moments is equivalent to the evaluation of certain sums or integrals. It is usually necessary to consider whether the moments exist. Some examples will illustrate the general principles involved.

Example 3.2

Consider the so-called binomial distribution $(q+p)^n$ in which the frequencies of values 0, h, $2h$, . . . are the successive terms in the expansion of the distribution, i.e. are $q^n, \binom{n}{1}q^{n-1}p, \binom{n}{2}q^{n-2}p^2$. Taking an origin at the first term and working in units of h, we have

$$\mu_1 \qquad \sum_{j=0} \qquad)q^{n-j}\, p^j\, j$$

which may be written

$$p\frac{\partial}{\partial p}(q+p)^n$$
$$= np(q+p)^{n-1}$$
$$= np.$$
$$\mu_2' = \sum_{j=0}^{n}\left\{\binom{n}{j}q^{n-j}\, p^j\, j^2\right.$$
$$= \left(p\frac{\partial}{\partial p}\right)^2(q+p)^n$$
$$= np(q+p)^{n-1} + n(n-1)p^2(q+p)^{n-2}$$
$$= n^2p^2 + npq.$$

Hence $\qquad \mu_2 = npq.$

Similarly $\qquad \mu_3' = \left(p\dfrac{\partial}{\partial p}\right)^3(q+p)^n$

etc., and it will be found that

$$\mu_3 = npq(q-p)$$
$$\mu_4 = 3p^2q^2n^2 + pqn(1-6pq).$$

Example 3.3

Consider the distribution

$$dF = \frac{k}{(1+x^2)^m}\,dx \qquad -\infty \leqslant x \leqslant \infty$$
$$m \geqslant 1.$$

This is a unimodal distribution symmetrical about $x = 0$. All existent moments of odd order about the origin therefore vanish. The constant k is given by the equation

$$1 = k\int_{-\infty}^{\infty} \frac{dx}{(1 + x^2)^m}$$
$$= k\frac{\Gamma(\frac{1}{2})\Gamma(m - \frac{1}{2})}{\Gamma(m)}.$$

The moment about the mean of order $2r$, if it exists, is given by

$$\mu_{2r} = k\int_{-\infty}^{\infty} \frac{x^{2r}}{(1 + x^2)^m}\, dx,$$

and this integral converges if and only if

$$2m > 2r + 1.$$

Thus the moments about the origin of order $< (2m - 1)$ exist and those of higher order do not.

If $m = 1$ it may be noted that the integral $\int_{-\infty}^{\infty} \frac{kx\, dx}{1 + x^2}$ is not completely convergent, i.e. $\lim_{\infty,\, n'} \int_{-n}^{n'} \frac{kx\, dx}{(1 + x^2)}$ does not exist, although the principal value

$$\lim_{n \to \infty} \int_{-n}^{n} \frac{kx\, dx}{(1 + x^2)}$$

does exist and is equal to zero. It is a matter of convention whether we regard the distribution as possessing a mean in this case. For $m > 1$ the mean exists and is located at the origin.

Making the substitution $z = \frac{1}{1 + x^2}$ in the formula for μ_{2r}, we find

$$\mu_{2r} = k\int_0^1 (1 - z)^{r-\frac{1}{2}} z^{m-r-\frac{3}{2}}\, dz$$
$$= k\frac{\Gamma(r + \frac{1}{2})\Gamma(m - r - \frac{1}{2})}{\Gamma(m)}$$

and on substituting for k,

$$\mu_{2r} = \frac{\Gamma(r + \frac{1}{2})\Gamma(m - r - \frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(m - \frac{1}{2})} \quad \text{if } 2m > 2r + 1.$$

Example 3.4

Consider the "normal" distribution

$$dF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\, dx \qquad -\infty \leqslant x \leqslant \infty.$$

This is symmetrical about the origin. All moments exist, those of odd order vanishing. Thus

$$\mu_{2r} = \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} x^{2r} e^{-\frac{x^2}{2\sigma^2}}\, dx.$$

This may be evaluated by partial integration, but a more direct method is as follows:

Consider the integral

$$M(t) = \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2\sigma^2}}\, dx$$

We have, for all real values of t

$$e^{tx}e^{-\frac{x^2}{2\sigma^2}} = \sum_{r=0}^{\infty}\left(\frac{t^r}{r!}\, x^r\, e^{-\frac{x^2}{2\sigma^2}}\right).$$

The series on the right is uniformly convergent in x and may be integrated term by term if the resulting series is uniformly convergent. We then have

$$M(t) = \sum_{r=0}^{\infty}\left(\frac{t^r}{r!}\mu_r\right).$$

In other words, μ_r is the coefficient of $\frac{t^r}{r!}$ in $e^{\frac{1}{2}\sigma^2 t^2}$ and hence

$$\mu_{2r} = \frac{\sigma^{2r}(2r)!}{2^r\ \ r!}.$$

Moment-generating Functions and Characteristic Functions

3.5. The previous example shows that in some cases we can derive from the distribution or the frequency function a function $M(t)$ which, when expanded in powers of t, will yield the moments of the distribution as the coefficients of those powers. Such a function is accordingly called a Moment-generating Function. It will be discussed more fully in the next chapter.

For many frequency functions the integral $\int_{-\infty}^{\infty} e^{tx}\,dF$ or the sum $\Sigma\{e^{tx_j}f(x_j)\}$ may not exist for real values of t. This is, for example, true of the function $dF = k(1 + x^2)^{-m}\,dx$ for finite positive values of m. A more serviceable auxiliary function is

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\,dF \quad (t \text{ real}) \quad . \quad . \quad . \quad . \quad . \quad (3.10)$$

This is known as the Characteristic Function and is of great theoretical importance. It will be seen in Chapter 4 that under certain general conditions the characteristic function determines and is completely determined by the distribution function. It also yields many valuable results in the theory of sampling.

Since by the nature of the distribution function the integral $\int_{-\infty}^{\infty} dF$ converges,

$$|\,\varphi(t)\,| \leqslant \int_{-\infty}^{\infty} |\,e^{itx}\,|\,dF \leqslant 1$$

and hence the Stieltjes integral (3.10) converges absolutely and uniformly in t. It may therefore be integrated under the summation signs with respect to t, and may be differentiated provided that the resulting expressions exist and are uniformly convergent. We have, for example, writing D_t for $\frac{d}{dt}$,

$$D_t^r\varphi(t) = i^r\int_{-\infty}^{\infty} e^{itx}x^r\,dF,$$

and hence, putting $t = 0$,

$$\mu'_r = (-i)^r[D_t^r\varphi(t)]_{t=0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.11)$$

provided that μ'_r exists. If $\varphi(t)$ be expanded in powers of t, μ'_r must thus be equal to the coefficient of $\frac{(it)^r}{r!}$ in the expansion. Thus the characteristic function is also a moment-generating function.

Example 3.5

Consider again the binomial $(q + p)^n$. Taking h as unit, we have

$$\varphi(t) = \sum_{j=0}^{n} \left\{ \binom{n}{j} q^{n-j}\, p^j\, e^{itj} \right\}$$
$$= (q + pe^{it})^n.$$

Hence
$$\mu_1' = (-i)\left[\frac{d}{dt}(q + pe^{it})^n\right]$$
$$= np$$
$$\mu_2' = (-i)^2\left[\frac{d^2}{dt^2}(q + pe^{it})^n\right]_{t=}$$
$$= np + n(n-1)p^2,$$

and so on.

Example 3.6

Consider the distribution

$$dF = \frac{a^\gamma}{\Gamma(\gamma)}\, x^{\gamma-1}\, e^{-ax} \qquad \begin{array}{l} a > 0,\ \gamma > 0 \\ 0 \leqslant x \leqslant \infty \end{array}$$

which is known as Pearson's Type III (cf. Chapter 6). The distribution may have a variety of shapes, depending on the value of γ, but moments of all orders exist in virtue of the convergence of the integral $\int_0^\infty x^r e^{-ax}\, dx$, the Γ-function integral. We have then, for the characteristic function,

$$\varphi(t) = \frac{a^\gamma}{\Gamma(\gamma)} \int_0^\infty e^{x(-a+it)}\, x^{\gamma-1}\, dx.$$

By the substitution $z = x(a - it)$ this becomes

$$\varphi(t) = \frac{a^\gamma}{\Gamma(\gamma)\,(a - it)^\gamma} \int_0^\infty e^{-z}\, z^{\gamma-1}\, dz$$

$$= \left(1 - \frac{it}{a}\right)^{-\gamma}$$

since $\int_0^\infty e^{-z} z^{\gamma-1}\, dz = \Gamma(\gamma)$ whether z is real or complex.

Hence
$$\varphi(t) = 1 + \gamma\frac{it}{a} + \frac{\gamma(\gamma+1)}{2!}\left(\frac{it}{a}\right)^2 + \ldots$$

and thus

$$\mu_1' = \frac{\gamma}{a}$$
$$\mu_2' = \frac{\gamma(\gamma+1)}{a^2}$$
$$\mu_3' = \frac{\gamma(\gamma+1)(\gamma+2)}{a^3}$$

and so on. In particular,

$$\mu_2 = \frac{\gamma}{a^2}$$
$$\mu_3 = \frac{2\gamma}{a^3}.$$

Absolute Moments

3.6. The quantity

$$\nu'_r = \int_{-\infty}^{\infty} |x - a|^r \, dF \qquad . \quad . \quad . \quad . \quad . \quad . \quad (3.12)$$

is called the absolute moment of order r about a. The absolute moments about the mean are written without primes.

If r is even, the absolute moment is clearly equal to the ordinary moment, and if the range of the distribution is positive the absolute moments about any point to the left of the start of the distribution are equal to the ordinary moments of corresponding order.

There are some interesting inequalities concerning the absolute moments. Referring to the function $E(t)$ of **2.7** and remembering that it is a non-decreasing function of t, we find, on putting $t = 0, 1, 2, \ldots$, that

$$(\nu'_1)^1 \leqslant (\nu'_2)^{\frac{1}{2}} \leqslant (\nu'_3)^{\frac{1}{3}} \ldots \leqslant (\nu'_r)^{\frac{1}{r}} \ldots \quad . \quad . \quad . \quad . \quad (3.13)$$

A more general inequality, due to Liapounoff, is

$$\nu_b'^{\,a-c} \leqslant \nu_c'^{\,a-b}\, \nu_a'^{\,b-c}, \; a \geqslant b \geqslant c \geqslant 0. \quad . \quad . \quad . \quad . \quad (3.14)$$

A proof of this result is sketched in Exercise 3.14. In particular, putting $b = \frac{1}{2}(a + c)$ we find

$$\nu'^2_{\frac{r+s}{2}} \leqslant \nu'_r \nu'_s \qquad . \quad (3.15)$$

Further, putting $a = 0$ in (3.14) we find

$$\nu_b'^{\,-c} \leqslant \nu_c'^{\,-b}$$

or

$$\nu_c'^{\frac{1}{c}} \leqslant \nu_b'^{\frac{1}{b}}, \; c \leqslant b$$

which is equivalent to (3.13).

Factorial Moments

3.7. The factorial expression

$$x(x - h)(x - 2h) \ldots (x - \overline{r - 1}h)$$

may conveniently be written $x^{[r]}$, a notation which brings out an analogy with the power x^r. Taking first differences with respect to x and with unit h, we have

$$\begin{aligned} \Delta x^{[r]} &= (x + h)^{[r]} - x^{[r]} \\ &= (x + h)x(x - h) \ldots (x - \overline{r - 2}h) - x(x - h) \qquad (x - \overline{r - 1}h) \\ &= rx^{[r-1]}h, \end{aligned}$$

which may be compared with the differential equation

$$dx^r = rx^{r-1}\,dx.$$

Conversely

$$\sum_{x=0}^{x} x^{[r]} = \frac{1}{(r + 1)h}(x + h)^{[r+1]}$$

corresponding to

$$\int_0^x x^r\,dx = \frac{1}{r + 1}x^{r+1}$$

The rth factorial moment about an arbitrary origin may then be defined by the equation

$$\mu'_{[r]} = \sum_{j=-\infty}^{\infty} (x_j - a)^{[r]} f(x_j) \quad . \quad . \quad . \quad . \quad . \quad (3.16)$$

where we have chosen the summation sign Σ rather than the Stieltjes integral because it is almost entirely for discontinuous distributions, or continuous distributions grouped in intervals of width h, that the factorial moments are used. In statistical theory they are not very prominent, but in the theory of interpolation and of curve fitting they are sufficiently important to justify some mention of their properties.

As usual, when it is necessary to distinguish between factorial moments about the mean and those about an arbitrary point we may write the former without the prime.

3.8. The factorial moments obey laws of transformation similar to those of equation (3.5) governing ordinary moments. In fact we have the expansion *

$$(a + b)^{[r]} = \sum_{j=0}^{r} \binom{r}{j} a^{[r-j]}\, b^{[j]}$$

and hence

$$(x - a)^{[r]} = (x - b + c)^{[r]} \text{ where } c = b - a$$

$$= \sum_{j=0}^{r} \binom{r}{j} (x - b)^{[r-j]}\, c^{[j]}$$

and hence

$$\mu'_{[r]}(a) = \sum_{j=0}^{r} \binom{r}{j} \mu'_{[r-j]}\, c^{[j]} \quad . \quad . \quad . \quad . \quad (3.17)$$

which may be written symbolically

$$\mu'_{[r]}(a) = (\mu'(b) + c)^{[r]}.$$

3.9. By direct expansion of (3.16) it is seen that

$$\begin{aligned} \mu'_{[1]} &= \mu'_1 \\ \mu'_{[2]} &= \mu'_2 - h\mu'_1 \\ \mu'_{[3]} &= \mu'_3 - 3h\mu'_2 + 2h^2\mu'_1 \\ \mu'_{[4]} &= \mu'_4 - 6h\mu'_3 + 11h^2\mu'_2 - 6h^3\mu'_1 \end{aligned} \quad . \quad (3.18)$$

and conversely that

$$\left.\begin{aligned} \mu'_2 &= \mu'_{[2]} + h\mu'_{[1]} \\ \mu'_3 &= \mu'_{[3]} + 3h\mu'_{[2]} + h^2\mu'_{[1]} \\ \mu'_4 &= \mu'_{[4]} + 6h\mu'_{[3]} + 7h^2\mu'_{[2]} + h^3\mu'_{[1]} \end{aligned}\right\} \quad . \quad (3.19)$$

Since the first moments are equal the equations remain true when the primes are dropped and terms in first moments omitted.

* It is clear that $(a + b)^{[r]}$ will be a polynomial of degree r in a, and may therefore be equated to $\sum_{j=0}^{r} k_j\, a^{[r-j]}$, where the k's are polynomials in b and h but do not contain a. Putting $a = 0$ we obtain $b^{[r]} = k_r$. Taking first differences with respect to a and putting $a = 0$ we obtain $rb^{[r-1]} = k_{r-1}$. Successive differences give the k's and the above result follows.

It is possible to give general formulae showing the factorial moments about one point in terms of the ordinary moments about another, and vice-versa. In fact

$$\mu_r'(a) = \sum_{j=0}^{r}\left\{\binom{r}{j}B_{r-j}^{(-j)}(c)h^{r-j}\mu_{[j]}'(b)\right. \qquad . \quad . \quad . \ (3.20)$$

$$\mu_{[r]}'(a) = \sum_{j=0}^{r}\left\{\binom{r}{j}B_{r-j}^{(r+1)}(c+h)h^{r-j}\mu_j'(b)\right. \qquad (3.21)$$

where $B_r^{(n)}(x)$ is the Bernouilli polynomial of order n and degree r in x, defined as the coefficient of $\frac{t^r}{r!}$ in $\left(\frac{t}{e^t-1}\right)^n e^{tx}$. For a discussion of these polynomials and the derivation of equations (3.20) and (3.21) reference may be made to Frisch (1926).

Calculation of Factorial Moments

3.10. The calculation of factorial moments for grouped data may be effected by a process of progressive summation which is illustrated in Table 3.2.

TABLE 3.2

(1) Frequency.	(2) First Summation.	(3) Second Summation.	(4) Third Summation.
f_1	$f_1 + \ldots + f_n$		
f_2	$f_2 + \ldots + f_n$	$f_2 + 2f_3 + \ldots (n-1)f_n$	
f_3	$f_3 + \ldots + f_n$	$f_3 + 2f_4 + \ldots (n-2)f_n$	$f_3 + 3f_4 + \ldots \frac{(n-1)(n-2)}{2}f_n$
f_4	$f_4 + f_5 + \ldots + f_n$	$f_4 + \ldots (n-3)f_n$	$f_4 + 3f_5 + \ldots \frac{(n-2)(n-3)}{2}f_n$
f_{n-2}	$f_{n-2} + f_{n-1} + f_n$	$f_{n-2} + 2f_{n-1} + 3f_n$	$f_{n-2} + 3f_{n-1} + 6f_n$
f_{n-1}	$f_{n-1} + f_n$	$f_{n-1} + 2f_n$	$f_{n-1} + 3f_n$
f_n	f_n	f_n	f_n
TOTALS	$S(jf_j) = \mu_{[1]}'$	$S\left\{\frac{j(j-1)}{2}f_j\right\} = \frac{1}{2!}\mu_{[2]}'$	$S\left\{\frac{j(j-1)(j-2)}{3!}f_j\right\} = \frac{1}{3!}\mu_{[3]}'$

Writing the proportional frequencies in the successive n intervals as $f_1 \ldots f_n$, as shown in the left-hand column, we construct column 2 by adding frequencies from the bottom. In the nth row we write f_n, in the $(n-1)$th row the sum $f_n + f_{n-1}$, in the $(n-2)$th row the sum $f_n + f_{n-1} + f_{n-2}$, and so on, the first row containing the sum $f_n + f_{n-1} + \ldots f_1$.

In column 3 the process is repeated with the rows of column 2, stopping at the second row, e.g. the nth row contains f_n, the $(n-1)$th row $(f_n + f_{n-1}) + f_n = 2f_n + f_{n-1}$, and so on, the second row containing the sum $(n-1)f_n + (n-2)f_{n-1} + \ldots + 2f_3 + f_2$.

Column 4 repeats the process with the entries of column 3, but stopping at the third row; and so on.

Consider now the sum of the entries in column 2. In that sum f_1 appears once, f_2 twice, . . . f_n n times. Hence the sum is equal to $\sum_{j=1}^{n}(jf_j)$

$$= \mu'_{[1]}.$$

In column 3, f_2 appears once, f_3 3 times, . . . f_n $\frac{1}{2}n(n-1)$ times. Hence

$$\text{Sum} = \sum_{j=1}^{n}\left\{\frac{j(j-1)}{2}f_j\right\}$$
$$= \tfrac{1}{2}\mu'_{[2]}.$$

In general, the sum of the $(r+1)$th column will be given by

$$\text{Sum} = \frac{1}{r!}\mu'_{[r]}.$$

If the actual frequencies are used instead of the proportional frequencies the sums have to be divided by the total frequency N.

Thus the process of summation gives the factorial moments directly. It is a modification of one which is due to G. F. Hardy (cf. Elderton, 1938*a*). The use of the method in practice lies in the fact that for certain calculating machines the progressive summation is easier to carry out than the processes involved in the method of Example 3.1.

Example 3.7

Consider again the data of Table 1.7, showing the distribution of 8585 men according to height in inches. The columns on the right in Table 3.3 overleaf show the successive sums. At the top of each column there has been placed within brackets the number which would have been obtained if the summation were continued up the column one place further than is required for the sum at the foot. These bracketed figures are useful to have as a check since each must equal the sum at the foot of the preceding column.

From this table we find

$$\begin{aligned}
\mu'_{[1]} &= 11{\cdot}020,850,320,33\\
\mu'_{[2]} &= 117{\cdot}055,096,097,84\\
\mu'_{[3]} &= 1,194{\cdot}957,483,983,69\\
\mu'_{[4]} &= 11,702{\cdot}727,082,119,98
\end{aligned}$$

From these values we may derive the ordinary moments, using equations (3.19), and find

$$\begin{aligned}
\mu'_1 &= 11{\cdot}020,850,320,33\\
\mu'_2 &= 128{\cdot}075,946,418,2\\
\mu'_3 &= 1,557{\cdot}143,622,597,5\\
\mu'_4 &= 19,702{\cdot}878,509,027,3,
\end{aligned}$$

from which we find, for the moments about the mean,

$$\begin{aligned}
\mu_2 &= 6{\cdot}616,805\\
\mu_3 &= -\,0{\cdot}207,840\\
\mu_4 &= 137{\cdot}689,185,
\end{aligned}$$

the units being one inch.

TABLE 3.3

Calculation of the Factorial Moments of a Distribution of Men according to Height in Inches (Table 1.7).

Height.	Frequency.	First Sum.	Second Sum.	Third Sum.	Fourth Sum.
57–	2	8,585	(94,614)	—	—
58–	4	8,583	86,029	(502,459)	—
59–	14	8,579	77,446	416,430	(1,709,785)
60–	41	8,565	68,867	338,984	1,293,355
61–	83	8,524	60,302	270,117	954,371
62–	169	8,441	51,778	209,815	684,254
63–	394	8,272	43,337	158,037	474,439
64–	669	7,878	35,065	114,700	316,402
65–	990	7,209	27,187	79,635	201,702
66–	1223	6,219	19,978	52,448	122,067
67–	1329	4,996	13,759	32,470	69,619
68–	1230	3,667	8,763	18,711	37,419
69–	1063	2,437	5,096	9,948	18,438
70–	646	1,374	2,659	4,852	8,490
71–	392	728	1,285	2,193	3,638
72–	202	336	557	908	1,445
73–	79	134	221	351	537
74–	32	55	87	130	186
75–	16	23	32	43	56
76–	5	7	9	11	13
77–	2	2	2	2	2
TOTALS	8585	94,614	502,459	1,709,785	4,186,163

Cumulants

3.11. The moments are a set of parameters of a distribution which are useful for measuring its properties and, in certain circumstances, for specifying it. Their use in these connections will be considered in later chapters. They are not, however, the only set of parameters for the purpose, or even the best set. Another series of parameters, the so-called cumulants, have properties which are more useful from the theoretical standpoint.

Formally, the cumulants $\kappa_1, \kappa_2, \ldots \kappa_r$ are defined by the identity in t

$$\exp\left\{\kappa t + \frac{\kappa_2 t^2}{2!} + \ldots + \frac{\kappa_r t^r}{r!} + \right.$$

$$= 1 + \mu_1' t + \frac{\mu_2' t^2}{2!} + \ldots + \frac{\mu_r' t^r}{r!} + \ldots \qquad (3.22)$$

It is sometimes more convenient to write the same equation with it for t, thus:

$$\exp\left\{\kappa(it) + \kappa_2\frac{(it)^2}{2!} + \qquad \kappa_r\frac{(it)^r}{r!} + \quad .\right\}$$

$$= 1 + \mu_1'\frac{(it)}{1!} + \qquad + \mu_r'\frac{(it)^r}{r!} +$$

$$= \int_{-\infty}^{\infty} \epsilon \quad dF$$

$$= \varphi(t) \qquad (3.23)$$

Thus, whereas μ'_r is the coefficient of $\frac{(it)^r}{r!}$ in $\varphi(t)$, the characteristic function, κ_r is the coefficient of $\frac{(it)^r}{r!}$ in $\log \varphi(t)$, if an expansion in power series exists.

3.12. If in equation (3.23) the origin is changed from a to b, where as usual $b - a = c$, the effect on $\varphi(t)$ is to multiply it by e^{-itc}, for $\int e^{itx}\,dF$ becomes $\int e^{it(x-c)}e^{itc}\,dF$. Hence the effect on $\log \varphi(t)$ is merely to add the term itc, and consequently the coefficients in $\log \varphi(t)$ are unchanged, except the first, which is decreased by c.

Hence the cumulants are invariant under change of origin, except the first. In this they stand in sharp contrast to the moments about an arbitrary point.

Both cumulants and moments have another property of an invariantive kind, namely, that if the variate-values are multiplied by a constant a, μ'_r and κ_r are multiplied by a^r. This is at once evident from their definitions. Thus any linear transformation of the kind

$$\xi = lx + m \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.24)$$

leaves the cumulants unchanged so far as the constant m is concerned and multiplies κ_r by l^r. The sole exception is the first cumulant which is equal to the mean. In particular, if we transform a distribution to standard measure, the only effect is to multiply κ_r by σ^{-r}, σ being the standard deviation and, as we shall see in a moment, being equal to $\kappa_2^{\frac{1}{2}}$.

The invariantive properties of the cumulants was the origin of their original name of semi-invariants, seminvariants or half-invariants (Thiele, 1889). It has, however, recently been shown that there are several other classes of parameters with the same property, and it seems best to reserve the word "seminvariant" for any parameter λ_r which, under the transformation (3.24), is multiplied by l^r. The cumulants and the moments *about the mean* are thus particular cases of seminvariants.

Relations between Moments and Cumulants

3.13. Subject to conditions of existence, we have, from (3.22)

$$\begin{aligned}1 + \mu'_1\frac{t}{1!} + \ldots + \mu'_r\frac{t_r}{r!} + \ldots \\ &= \exp\left\{\kappa_1\frac{t}{1!} + \kappa_2\frac{t^2}{2!} + \quad \kappa_r\frac{}{r!}\ldots\right. \\ &= \exp\left(\frac{\kappa_1 t}{1!}\right)\exp\left(\frac{\kappa_2 t^2}{2!}\right)\ldots\exp\left(\frac{\kappa_r t^r}{r!}\right)\ldots \\ &= \left(1 + \frac{\kappa_1 t}{1!} + \frac{\kappa_1{}^2 t^2}{2!} \quad \right. \left\{1 + \frac{\kappa_2 t^2}{2!} + \frac{1}{2!}\left(\frac{\kappa_2^2 t^4}{2!^2}\right) + \ldots\right. \\ &\qquad 1 + \left(\frac{\kappa_r t^r}{r!}\right) + \frac{1}{2!}\left(\frac{\kappa_r t^r}{r!}\right)^2\end{aligned}$$

Picking out the terms in the exponential expansions which, when multiplied together, give a power of t^r, we have

$$\mu'_r = \sum_{m=0}^{r}\sum\left(\frac{\kappa_{p_1}}{p_1!}\right)^{\pi_1}\left(\frac{\kappa_{p_2}}{p_2!}\right)^{\pi_2}\ldots\left(\frac{\kappa_{p_m}}{p_m!}\right)^{\pi_m}\frac{r!}{\pi_1!\pi_2!\ldots\pi_m!}. \quad . \quad . \quad (3.25)$$

where the second summation extends over all non-negative values of the π's such that

$$p_1\pi_1 + p_2\pi_2 + \ldots p_m\pi_m = r. \qquad (3.26)$$

It is worth noting that the rather tedious process of writing down the explicit relations for particular values of r may be shortened considerably. In fact, differentiating (3.22) by κ_j we have

$$\frac{t^j}{j!}\left(1 + \mu_1' t + \ldots + \frac{\mu_r' t^r}{r!} + \ldots\right) = \frac{\partial \mu_1'}{\partial \kappa_j} + \ldots + \frac{t^r}{r!}\frac{\partial \mu'_r}{\partial \kappa_j} + \ldots$$

and hence, identifying powers of t,

$$\frac{\partial \mu_r'}{\partial \kappa_j} = \binom{r}{j}\mu'_{r-j} \qquad (3.27)$$

In particular

$$\frac{\partial \mu_r'}{\partial \kappa_1} = r\mu'_{r-1} \qquad (3.28)$$

and thus, given any μ_r' in terms of the κ's we can write down successively those of lower orders by a differentiation.

The first ten of these expressions are, for moments about an arbitrary point :—

$$\begin{aligned}
\mu_1' &= \kappa_1, \\
\mu_2' &= \kappa_2 + \kappa_1^2, \\
\mu_3' &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3, \\
\mu_4' &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4, \\
\mu_5' &= \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3 + \kappa_1^5, \\
\mu_6' &= \kappa_6 + 6\kappa_5\kappa_1 + 15\kappa_4\kappa_2 + 15\kappa_4\kappa_1^2 + 10\kappa_3^2 + 60\kappa_3\kappa_2\kappa_1 + 20\kappa_3\kappa_1^3 \\
&\quad + 15\kappa_2^3 + 45\kappa_2^2\kappa_1^2 + 15\kappa_2\kappa_1^4 + \kappa_1^6, \\
\mu_7' &= \kappa_7 + 7\kappa_6\kappa_1 + 21\kappa_5\kappa_2 + 21\kappa_5\kappa_1^2 + 35\kappa_4\kappa_3 + 105\kappa_4\kappa_2\kappa_1 \\
&\quad + 35\kappa_4\kappa_1^3 + 70\kappa_3^2\kappa_1 + 105\kappa_3\kappa_2^2 + 210\kappa_3\kappa_2\kappa_1^2 + 35\kappa_3\kappa_1^4 \\
&\quad + 105\kappa_2^3\kappa_1 + 105\kappa_2^2\kappa_1^3 + 21\kappa_2\kappa_1^5 + \kappa_1^7, \\
\mu_8' &= \kappa_8 + 8\kappa_7\kappa_1 + 28\kappa_6\kappa_2 + 28\kappa_6\kappa_1^2 + 56\kappa_5\kappa_3 + 168\kappa_5\kappa_2\kappa_1 + 56\kappa_5\kappa_1^3 \\
&\quad + 35\kappa_4^2 + 280\kappa_4\kappa_3\kappa_1 + 210\kappa_4\kappa_2^2 + 420\kappa_4\kappa_2\kappa_1^2 + 70\kappa_4\kappa_1^4 \\
&\quad + 280\kappa_3^2\kappa_2 + 280\kappa_3^2\kappa_1^2 + 840\kappa_3\kappa_2^2\kappa_1 + 560\kappa_3\kappa_2\kappa_1^3 + 56\kappa_3\kappa_1^5 \\
&\quad + 105\kappa_2^4 + 420\kappa_2^3\kappa_1^2 + 210\kappa_2^2\kappa_1^4 + 28\kappa_2\kappa_1^6 + \kappa_1^8, \\
\mu_9' &= \kappa_9 + 9\kappa_8\kappa_1 + 36\kappa_7\kappa_2 + 36\kappa_7\kappa_1^2 + 84\kappa_6\kappa_3 + 252\kappa_6\kappa_2\kappa_1 \\
&\quad + 84\kappa_6\kappa_1^3 + 126\kappa_5\kappa_4 + 504\kappa_5\kappa_3\kappa_1 + 378\kappa_5\kappa_2^2 + 756\kappa_5\kappa_2\kappa_1^2 \\
&\quad + 126\kappa_5\kappa_1^4 + 315\kappa_4^2\kappa_1 + 1260\kappa_4\kappa_3\kappa_2 + 1260\kappa_4\kappa_3\kappa_1^2 + 1890\kappa_4\kappa_2^2\kappa_1 \\
&\quad + 1260\kappa_4\kappa_2\kappa_1^3 + 126\kappa_4\kappa_1^5 + 280\kappa_3^3 + 2520\kappa_3^2\kappa_2\kappa_1 + 840\kappa_3^2\kappa_1^3 \\
&\quad + 1260\kappa_3\kappa_2^3 + 3780\kappa_3\kappa_2^2\kappa_1^2 + 1260\kappa_3\kappa_2\kappa_1^4 + 84\kappa_3\kappa_1^6 + 945\kappa_2^4\kappa_1 \\
&\quad + 1260\kappa_2^3\kappa_1^3 + 378\kappa_2^2\kappa_1^5 + 36\kappa_2\kappa_1^7 + \kappa_1^9, \\
\mu_{10}' &= \kappa_{10} + 10\kappa_9\kappa_1 + 45\kappa_8\kappa_2 + 45\kappa_8\kappa_1^2 + 120\kappa_7\kappa_3 + 360\kappa_7\kappa_2\kappa_1 \\
&\quad + 120\kappa_7\kappa_1^3 + 210\kappa_6\kappa_4 + 840\kappa_6\kappa_3\kappa_1 + 630\kappa_6\kappa_2^2 + 1260\kappa_6\kappa_2\kappa_1^2 \\
&\quad + 210\kappa_6\kappa_1^4 + 126\kappa_5^2 + 1260\kappa_5\kappa_4\kappa_1 + 2520\kappa_5\kappa_3\kappa_2 + 2520\kappa_5\kappa_3\kappa_1^2 \\
&\quad + 3780\kappa_5\kappa_2^2\kappa_1 + 2520\kappa_5\kappa_2\kappa_1^3 + 252\kappa_5\kappa_1^5 + 1575\kappa_4^2\kappa_2 + 1575\kappa_4^2\kappa_1^2 \\
&\quad + 2100\kappa_4\kappa_3^2 + 12600\kappa_4\kappa_3\kappa_2\kappa_1 + 4200\kappa_4\kappa_3\kappa_1^3 + 3150\kappa_4\kappa_2^3 \\
&\quad + 9450\kappa_4\kappa_2^2\kappa_1^2 + 3150\kappa_4\kappa_2\kappa_1^4 + 210\kappa_4\kappa_1^6 + 2800\kappa_3^3\kappa_1 \\
&\quad + 6300\kappa_3^2\kappa_2^2 + 12600\kappa_3^2\kappa_2\kappa_1^2 + 2100\kappa_3^2\kappa_1^4 + 12600\kappa_3\kappa_2^3\kappa_1 \\
&\quad + 12600\kappa_3\kappa_2^2\kappa_1^3 + 2520\kappa_3\kappa_2\kappa_1^5 + 120\kappa_3\kappa_1^7 + 945\kappa_2^5 + 4725\kappa_2^4\kappa_1^2 \\
&\quad + 3150\kappa_2^3\kappa_1^4 + 630\kappa_2^2\kappa_1^6 + 45\kappa_2\kappa_1^8 + \kappa_1^{10} ;
\end{aligned} \qquad (3.29)$$

or, for moments about the mean ($\kappa_1 = 0$),

$$
\begin{aligned}
\mu_2 &= \kappa_2, \\
\mu_3 &= \kappa_3, \\
\mu_4 &= \kappa_4 + 3\kappa_2^2, \\
\mu_5 &= \kappa_5 + 10\kappa_3\kappa_2, \\
\mu_6 &= \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3, \\
\mu_7 &= \kappa_7 + 21\kappa_5\kappa_2 + 35\kappa_4\kappa_3 + 105\kappa_3\kappa_2^2, \\
\mu_8 &= \kappa_8 + 28\kappa_6\kappa_2 + 56\kappa_5\kappa_3 + 35\kappa_4^2 + 210\kappa_4\kappa_2^2 + 280\kappa_3^2\kappa_2 + 105\kappa_2^4, \\
\mu_9 &= \kappa_9 + 36\kappa_7\kappa_2 + 84\kappa_6\kappa_3 + 126\kappa_5\kappa_4 + 378\kappa_5\kappa_2^2 + 1260\kappa_4\kappa_3\kappa_2 + 280\kappa_3^3 \\
&\quad + 1260\kappa_3\kappa_2^3, \\
\mu_{10} &= \kappa_{10} + 45\kappa_8\kappa_2 + 120\kappa_7\kappa_3 + 210\kappa_6\kappa_4 + 630\kappa_6\kappa_2^3 + 126\kappa_5^2 \\
&\quad + 2520\kappa_5\kappa_3\kappa_2 + 1575\kappa_4^2\kappa_2 + 2100\kappa_4\kappa_3^2 + 3150\kappa_4\kappa_2^3 \\
&\quad + 6300\kappa_3^2\kappa_2^2 + 945\kappa_2^3.
\end{aligned}
\tag{3.30}
$$

Conversely we have

$$\frac{\kappa_1 t}{1!} + \dots + \frac{\kappa_r t^r}{r!} + \dots = \log\left(1 + \frac{\mu_1' t}{1!} + \dots \frac{\mu_r' t^r}{r!} + \dots\right).$$

Expanding the logarithm and picking out powers of t^r as before, we have

$$\kappa_r = r! \sum_{m=0}^{r} \sum \left(\frac{\mu_{p_1}'}{p_1!}\right)^{\pi_1} \dots \left(\frac{\mu_{p_m}'}{p_m!}\right)^{\pi_m} \frac{(-1)^{\rho-1}(\rho-1)!}{\pi_1! \dots \pi_m!} \quad . \tag{3.31}$$

the second summation extending over all non-negative π's and ρ's, subject to (3.26) and the further condition

$$\pi_1 + \pi_2 + \dots \pi_m = \rho \tag{3.32}$$

The first ten formulae are, in terms of moments about an arbitrary point:—

$$
\begin{aligned}
\kappa_1 &= \mu_1', \\
\kappa_2 &= \mu_2' - \mu_1'^2, \\
\kappa_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3, \\
\kappa_4 &= \mu_4' - 4\mu_3'\mu_1' - 3\mu_2'^2 + 12\mu_2'\mu_1'^2 - 6\mu_1'^4, \\
\kappa_5 &= \mu_5' - 5\mu_4'\mu_1' - 10\mu_3'\mu_2' + 20\mu_3'\mu_1'^2 + 30\mu_2'^2\mu_1' - 60\mu_2'\mu_1'^3 + 24\mu_1'^5, \\
\kappa_6 &= \mu_6' - 6\mu_5'\mu_1' - 15\mu_4'\mu_2' + 30\mu_4'\mu_1'^2 - 10\mu_2'^3 + 120\mu_3'\mu_2'\mu_1' - 120\mu_3'\mu_1'^3 \\
&\quad + 30\mu_2'^3 - 270\mu_2'^2\mu_1'^2 + 360\mu_2'\mu_1'^4 - 120\mu_1'^6, \\
\kappa_7 &= \mu_7' - 7\mu_6'\mu_1' - 21\mu_5'\mu_2' + 42\mu_5'\mu_1'^2 - 35\mu_4'\mu_3' + 210\mu_4'\mu_2'\mu_1' \\
&\quad - 210\mu_4'\mu_1'^3 + 140\mu_3'^2\mu_1' + 210\mu_3'\mu_2'^2 - 1260\mu_3'\mu_2'\mu_1'^2 + 840\mu_3'\mu_1'^4 \\
&\quad - 630\mu_2'^3\mu_1' + 2520\mu_2'^2\mu_1'^3 - 2520\mu_2'\mu_1'^5 + 720\mu_1'^7, \\
\kappa_8 &= \mu_8' - 8\mu_7'\mu_1' - 28\mu_6'\mu_2' + 56\mu_6'\mu_1'^2 - 56\mu_5'\mu_3' + 336\mu_5'\mu_2'\mu_1' \\
&\quad - 336\mu_5'\mu_1'^3 - 35\mu_4'^2 + 560\mu_4'\mu_3'\mu_1' + 420\mu_4'\mu_2'^2 - 2520\mu_4'\mu_2'\mu_1'^2 \\
&\quad + 1680\mu_4'\mu_1'^4 + 560\mu_3'^2\mu_2' - 1680\mu_3'^2\mu_1'^2 - 5040\mu_3'\mu_2'^2\mu_1' \\
&\quad + 13440\mu_3'\mu_2'\mu_1'^3 - 6720\mu_3'\mu_1'^5 - 630\mu_2'^4 + 10080\mu_2'^3\mu_1'^2 \\
&\quad - 25200\mu_2'^2\mu_1'^4 + 20160\mu_2'\mu_1'^6 - 5040\mu_1'^8, \\
\kappa_9 &= \mu_9' - 9\mu_8'\mu_1' - 36\mu_7'\mu_2' + 72\mu_7'\mu_1'^2 - 84\mu_6'\mu_3' + 504\mu_6'\mu_2'\mu_1' \\
&\quad - 504\mu_6'\mu_1'^3 - 126\mu_5'\mu_4' + 1008\mu_5'\mu_3'\mu_1' + 756\mu_5'\mu_2'^2 - 4536\mu_5'\mu_2'\mu_1'^2 \\
&\quad + 3024\mu_5'\mu_1'^4 + 630\mu_4'^2\mu_1' + 2520\mu_4'\mu_3'\mu_2' - 7560\mu_4'\mu_3'\mu_1'^2 \\
&\quad - 11340\mu_4'\mu_2'^2\mu_1' + 30240\mu_4'\mu_2'\mu_1'^3 - 15120\mu_4'\mu_1'^5 + 560\mu_3'^3 \\
&\quad - 15120\mu_3'^2\mu_2'\mu_1' + 20160\mu_3'^2\mu_1'^3 - 7560\mu_3'\mu_2'^2 + 90720\mu_3'\mu_2'^2\mu_1'^2 \\
&\quad - 151200\mu_3'\mu_2'\mu_1'^4 + 60480\mu_3'\mu_1'^6 + 22680\mu_2'^4\mu_1' - 151200\mu_2'^3\mu_1'^3 \\
&\quad + 272160\mu_2'^2\mu_1'^5 - 181440\mu_2'\mu_1'^7 + 40320\mu_1'^9,
\end{aligned}
\tag{3.33}
$$

$$
\begin{aligned}
\kappa_{10} = \mu'_{10} &- 10\mu'_9\mu'_1 - 45\mu'_8\mu'_2 + 90\mu'_8\mu'^2_1 - 120\mu'_7\mu'_3 + 720\mu'_7\mu'_2\mu'_1 \\
&- 720\,\mu'_7\mu'^3_1 - 210\mu'_6\mu'_4 + 1680\mu'_6\mu'_3\mu'_1 + 1260\mu'_6\mu'^2_2 \\
&- 7560\mu'_6\mu'_2\mu'^2_1 + 5040\mu'_6\mu'^4_1 - 126\mu'^2_5 + 2520\mu'_5\mu'_4\mu'_1 \\
&+ 5040\mu'_5\mu'_3\mu'_2 - 15120\mu'_5\mu'_3\mu'^2_1 - 22680\mu'_5\mu'^2_2\mu'_1 + 60480\mu'_5\mu'_2\mu'^3_1 \\
&- 30240\mu'_5\mu'^5_1 + 3150\mu'^2_4\mu'_2 - 9450\mu'^2_4\mu'^2_1 + 4200\mu'_4\mu'^2_3 \\
&- 75600\mu'_4\mu'_3\mu'_2\mu'_1 + 100800\mu'_4\mu'_3\mu'^3_1 - 18900\mu'_4\mu'^3_2 \\
&+ 226800\mu'_4\mu'^2_2\mu'^2_1 - 378000\mu'_4\mu'_2\mu'^4_1 + 151200\mu'_4\mu'^6_1 - 16800\mu'^3_3\mu'_1 \\
&- 37800\mu'^2_3\mu'^2_2 + 302400\mu'^2_3\mu'_2\mu'^2_1 - 252000\mu'^2_3\mu'^4_1 + 302400\mu'_3\mu'^3_2\mu'_1 \\
&- 1512000\mu'_3\mu'^2_2\mu'^3_1 + 1814400\mu'_3\mu'_2\mu'^5_1 - 604800\mu'_3\mu'^7_1 \\
&+ 22680\mu'^5_2 - 567000\mu'^4_2\mu'^2_1 + 2268000\mu'^3_2\mu'^4_1 - 3175200\mu'^2_2\mu'^6_1 \\
&+ 1814400\mu'_2\mu'^8_1 - 362880\mu'^{10}_1.
\end{aligned}
\tag{3.33}
$$

or, for moments about the mean,

$$
\begin{aligned}
\kappa_2 &= \mu_2, \\
\kappa_3 &= \mu_3, \\
\kappa_4 &= \mu_4 - 3\mu_2^2, \\
\kappa_5 &= \mu_5 - 10\mu_3\mu_2, \\
\kappa_6 &= \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3, \\
\kappa_7 &= \mu_7 - 21\mu_5\mu_2 - 35\mu_4\mu_3 + 210\mu_3\mu_2^2, \\
\kappa_8 &= \mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 + 560\mu_3^2\mu_2 - 630\mu_2^4, \\
\kappa_9 &= \mu_9 - 36\mu_7\mu_2 - 84\mu_6\mu_3 - 126\mu_5\mu_4 + 756\mu_5\mu_2^2 + 2520\mu_4\mu_3\mu_2 \\
&\quad + 560\mu_3^3 - 7560\mu_3\mu_2^3, \\
\kappa_{10} &= \mu_{10} - 45\mu_8\mu_2 - 120\mu_7\mu_3 - 210\mu_6\mu_4 + 1260\mu_6\mu_2^2 - 126\mu_5^2 \\
&\quad + 5040\mu_5\mu_3\mu_2 + 3150\mu_4^2\mu_2 + 4200\mu_4\mu_3^2 - 18900\mu_4\mu_2^3 \\
&\quad - 37800\mu_3^2\mu_2^2 + 22680\mu_2^5.
\end{aligned}
\tag{3.34}
$$

Existence of Cumulants

3.14. The formal expression (3.22) may be regarded as defining the cumulants in terms of the moments, and it is thus evident that the cumulant of order r exists if the moments of orders r and lower exist. If, however, we look to the equation

$$\exp\left(\Sigma\kappa_r\frac{(it)^r}{r!}\right) \qquad \phi(t)$$

as defining the cumulants, it is not quite so easy to show that κ_r exists if μ_r and lower μ's exist. It may, however, be shown that κ_r exists if ν_r, the absolute moment, exists, and this is sufficient for all ordinary purposes. Some care is necessary with the proof because the variable t in the characteristic function is real, but there also appears the complex quantity i.

We have

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF.$$

Expanding the exponential we have, if the moments up to μ_r exist,

$$\phi(t) = \sum_{j=0} \frac{\mu_j}{j!}(it)^j + R_r,$$

where

$$R_r = \int_{-\infty}^{\infty} dF\left(e^{ixt} - \sum_{j=0}^{r}\frac{(itx)^j}{j!}\right)$$

$$= \int_{-\infty}^{\infty} dF\left(\cos xt + i\sin xt - \sum_{j=0}^{r}\frac{(itx)^j}{j!}\right).$$

Considering the real and imaginary terms separately, we have, if r is even—

$$R_r = \int_{-\infty}^{\infty} dF\left(\cos xt - \sum_{j=0}^{\frac{r}{2}}(-1)^j\frac{(xt)^{2j}}{(2j)!}\right) + i\int_{-\infty}^{\infty} dF\left(\sin xt - \sum_{j=1}^{\frac{r}{2}}(-1)^{j-1}\frac{(xt)^{2j-1}}{(2j-1)!}\right).$$

The real term in the integrand consists of $(-1)^{\frac{r}{2}}\frac{(xt)^r}{r!}$ plus $\cos xt$ minus the first $\frac{r}{2}$ terms of the Maclaurin expansion of $\cos xt$, and is thus equal to $\frac{(xt)^r}{r!}\left(\frac{d^{\frac{r}{2}}}{dx^{\frac{r}{2}}}\cos xt\right)_{xt\,=\,\theta,}$ where $0 \leqslant \theta \leqslant 1$. The modulus of the term is thus not greater than $2\frac{|xt|^r}{r!}$. Similarly for the imaginary terms. Hence

$$|R_r| \leqslant 3\int_{-\infty}^{\infty}\frac{|xt|^r}{r!}dF$$

$$\leqslant 3\nu_r\frac{x^r}{r!}.$$

A similar result follows if r is odd. Now if μ'_r exists, it does not necessarily follow that ν'_r exists. But if the latter exists we have

$$\phi(t) = \sum_{j=0}^{\infty}\frac{\mu'_j(it)^j}{j!} + 0(t^r).$$

We may then, for some small t, take logarithms and expand, obtaining

$$\log\phi(t) = \sum_{j=0}^{\infty}\kappa_j\frac{(it)^j}{j!} + 0(t^r) \qquad (3.35)$$

the coefficients κ_j being the cumulants by definition. Hence if ν_r exists, κ_r and cumulants of lower orders exist.

Calculation of Cumulants

3.15. The cumulants are not, like the moments, determinable directly by summatory or integrative processes, and to find them it is necessary either to ascertain the moments and then apply equations (3.33), or to derive them from the characteristic function. For the latter case we have, from (3.35)

$$\kappa_r = (-i)^r\left[D_t^r\log\phi(t)\right]_{t=0} \qquad (3.36)$$

The following examples will illustrate the processes involved.

Example 3.8

In Example 3.7 we found the following values for the moments about the mean of the height data of Table 1.7 :—

$$\begin{aligned}\mu_1' &= 11{\cdot}020{,}850\\ \mu_2 &= 6{\cdot}616{,}805\\ \mu_3 &= -\ 0{\cdot}207{,}840\\ \mu_4 &= 137{\cdot}689{,}185,\end{aligned}$$

whence, from (3.34) κ_2 and κ_3 have the same values as μ_2 and μ_3 and

$$\begin{aligned}\kappa_4 &= \mu_4 - 3\mu_2^2\\ &= 6{\cdot}342{,}86.\end{aligned}$$

κ_1 is the same as μ_1', in this case measured from the centre of the interval 56– inches.

The same results would, of course, have been obtained if we had used equations (3.33) and moments about the origin.

Example 3.9

Consider the discontinuous distribution whose frequencies at the values $0, 1, \ldots j \ldots$ are $e^{-m}\left(1, \frac{m}{1!}, \ldots \frac{m^j}{j!} \ldots\right)$. The characteristic function is given by

$$\begin{aligned}\phi(t) &= e^{-m} \sum_{j=0}^{\infty} \frac{m^j}{j!} e^{ijt}\\ &= e^{-m} \exp\ (me^{it})\\ &= \exp\ m(e^{it} - 1).\end{aligned}$$

Since for any r the absolute moment is the same as the ordinary moment, we have

$$\mu_r' = e^{-m} \sum_{j=0}^{\infty} \frac{m^j j^r}{j!},$$

and since this converges * cumulants of all orders exist. They are therefore given by the expansion of $\log \phi(t)$ as a power series in t. But

$$\begin{aligned}\log \phi(t) &= m(e^{it} - 1)\\ &= m \sum_{j=1}^{\infty} \frac{(it)^j}{j!}\end{aligned}$$

and hence
$$\kappa_r = m$$
for all r. Thus all cumulants of the distribution are equal to m.

* For the ratio of the $(n + 1)$th term of the series to the nth is

$$\frac{m^{n+1}(n+1)^r}{(n+1)!} \Big/ \frac{m^n n^r}{\cdot} \qquad m \qquad 1 + \text{-}$$

and thus the series converges for all finite values of m.

Example 3.10

In Example 3.4 we found, in effect, for the characteristic function of the normal distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$\phi(t) = \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{ixt}e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= e^{-\frac{t^2\sigma^2}{2}}$$

$$\log \phi(t) = -\frac{t^2\sigma^2}{2}.$$

It is easily seen that the absolute moments and hence cumulants of all orders exist. Thus κ_r is the coefficient of $\frac{(it)^r}{r!}$ in $\log \phi(t)$, i.e. for the normal distribution all cumulants of order higher than the second are zero. The second cumulant is equal to σ^2.

Example 3.11

In Example 3.6 it was found that for the distribution

$$dF = \frac{a^\gamma}{\Gamma(\gamma)} x^{\gamma-1}e^{-ax} \qquad \begin{matrix} a > 0, \ \gamma > 0 \\ 0 \leqslant x \leqslant \infty \end{matrix}$$

the characteristic function is given by

$$\phi(t) = \frac{1}{\left(1 - \frac{it}{a}\right)^\gamma}$$

It is readily verified that cumulants of all orders exist and hence

$$\kappa^r = \text{coeff. of } \frac{(it)^r}{r!} \text{ in } -\gamma \log\left(1 - \frac{it}{a}\right)$$

$$= \gamma(r-1)!\, a^{-r}.$$

Example 3.12

Consider again the distribution of Example 3.3.

$$dF = \frac{k}{(1+x^2)^m} dx \qquad m \geqslant 1, \ -\infty \leqslant x \leqslant \infty.$$

The characteristic function is given by

$$\phi(t) = k\int_{-\infty}^{\infty} \frac{e^{ixt}}{(1+x^2)^m} dx,$$

which, since $\sin xt$ is an odd function, reduces to

$$k\int_{-\infty}^{\infty} \frac{\cos xt}{(1+x^2)^m} dx.$$

This integral may be evaluated by complex integration round a contour consisting of the

x-axis, the infinite semicircle above the x-axis and the infinitely small circle round the point $x = i$. It is found*

$$\phi(t) = \frac{k\pi}{2^{2m-2}(m-1)!}e^{-|t|}\Big\{(2|t|)^{m-1} + m(m-1)(2|t|)^{m-2}$$
$$+ \frac{(m+1)(m)(m-1)(m-2)}{2!}(2|t|)^{m-3} + \ldots + \frac{(2m-2)!}{(m-1)!}\Big\}.$$

If $r < 2m - 1$ the absolute moment of order r

$$\nu_r = k\int_{-\infty}^{\infty}\frac{|x|^r\,dx}{(1+x^2)^m}$$

exists and hence so does the cumulant of order r. But in this case we cannot expand $\log \phi(t)$ in an *infinite* series of powers of t, though this might perhaps be thought possible from the form of $\phi(t)$. In fact, we can only expand $\log \phi(t)$ in powers of t up to the point at which the differential coefficients of $\phi(t)$ exist, for $t = 0$.

To simplify the discussion, consider the case when $m = 2$. We have then, since $k = 2/\pi$ in this case,

$$\phi(t) = e^{-|t|}\{|t| + 1\}$$
$$\log \phi(t) = -|t| + \log\{1 + |t|\}$$

If t is positive this equals

$$-\frac{t^2}{2} + \frac{t^3}{3} - \ldots$$

but if t is negative it equals

$$-\frac{t^2}{2} - \frac{t^3}{3} - \ldots$$

the two expressions differing in the sign of the term in t^3 and every second term thereafter. There is thus no unique expansion of $\log \phi(t)$ in powers of t about the point $t = 0$. There are two forms of the function expressing $\log \phi(t)$ according as t is positive or negative.

However, these expressions coincide as far as their terms in t and t^2, and the first and second differential coefficients of $\log \phi(t)$ are uniquely defined when $t = 0$. Thus the first and second cumulants exist and are given by

$$\kappa_1 = 0$$
$$\kappa_2 = 1.$$

Cumulants of higher orders do not exist.

Corrections for Grouping

3.16. When moments are calculated from a numerically specified distribution which is grouped, there is present a certain amount of approximation owing to the fact that

* Results of this kind are given in several text-books of analysis, sometimes incorrectly, e.g. it is sometimes stated that

$$\int_{-\infty}^{\infty}\frac{\cos tx}{1+x^2}\,dx = \pi e^{-t},$$

which is only true when $t \geqslant 0$. The appearance of the modulus in the expression above is crucial for the purposes of the example. A correct proof is given in J. Edwards, *Integral Calculus*, vol. 2, article 1326.

the frequencies are assumed to be concentrated at the mid-points of intervals. It is possible to correct for this effect under certain conditions.

Suppose the frequency function $f(x)$ to be continuous. If the range is divided into intervals of width h, we are given, not the values of $f(x)$ at all points but the frequencies in those intervals, e.g. the frequency in the jth interval, centred at x_j, will be

$$f_j = \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j + \xi)\, d\xi.$$

We will denote the moments calculated from grouped frequencies—the "raw" moments—with a bar, so that we have

$$\begin{aligned}\bar{\mu}'_r &= \sum_{j=-\infty}^{\infty} (x_j{}^r f_j)\\ &= \sum_{j=-\infty}^{\infty} \left\{x_j{}^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j + \xi)\, d\xi\right\}.\end{aligned} \quad . \quad (3.37)$$

The true moment, if it exists, is given by

$$\mu'_r = \int_{-\infty}^{\infty} x^r f(x)\, dx$$

and it is required to investigate the relationship between the $\bar{\mu}$'s and the μ's.

Now we have, in virtue of the Euler-Maclaurin sum formula, for an arbitrary function $\kappa(x)$ which has derivatives of the mth order,

$$\frac{1}{h}\int_a^{a+nh} \kappa(x)\, dx = \{\tfrac{1}{2}\kappa(a) + \kappa(a+h) + \kappa(a+2h) + \ldots + \kappa(a + \overline{n-1}h) + \tfrac{1}{2}\kappa(a+nh)\}$$

$$\sum_{j=2}^{m-1} \frac{h^{j-1}}{j!} B_j \left[\kappa^{(j-1)}(x)\right]_a^{a+nh} - S_m \quad . \quad (3.38)$$

where S_m is a remainder term which may be expressed as

$$S_m = -\frac{nh^m}{m!} B_m \kappa^{(m)}(a + \theta nh) \qquad 0 \leqslant \theta \leqslant 1$$

m even,

and

$$S_m = \frac{2nh^m}{m!} B^{(1)}_{m+1}(\tfrac{1}{2}) \kappa^{(m)}(a + \theta nh), \qquad 0 \leqslant \theta \leqslant 1$$

m odd.*

Suppose now that $f(x)$ is of finite range, from a to b, derivable up to the mth order,

* Cf. Milne Thomson, *The Calculus of Finite Differences*, section 7.5, for the general Euler-Maclaurin expansion. The form of S_m when m is even is given in section 7.5 of that book, and the above form when m is odd may be derived similarly.

In our convention the Bernouilli number B_j is defined as the coefficient of $t^j/j!$ in $t/(e^t - 1)$. The Bernouilli polynomial has already been defined in 3.9. Explicitly $B_0 = 1$, $B_1 = \frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_3 = B_5 = B_{2j+1} = 0$, $B_4 = -\frac{1}{30}$, $B_6 = \frac{1}{42}$, $B_8 = -\frac{1}{30}$, $B_{10} = -\frac{5}{66}$, $B_{12} = \frac{691}{2730}$, $B_{14} = -\frac{7}{6}$.

and that at the end of the range $f(x)$ and its first m derivatives vanish. Then $f(x)$ and the first m derivatives are continuous throughout the range $-\infty$ to $+\infty$ and the function

$$\kappa(x) = x^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+\xi)\, d\xi \quad . \quad . \tag{3.39}$$

together with its first $m + 1$ derivatives, will also be continuous throughout that range. If a is infinite (and similarly for b) it is assumed that

$$\lim_{x \to -\infty} x^r f^{(j)}(x) \to 0$$

for all values of j up to and including m, in which case $\kappa(x)$ and its first $m + 1$ derivatives will also tend to zero. Thus in either case the Euler-Maclaurin expansion (3.38) is valid for $\kappa(x)$ given by (3.39) and we may write

$$\Big[\kappa^{(j)}(x)\Big]_{-\infty}^{\infty} = 0 \qquad j \leqslant m + 1.$$

Substituting in (3.38) we have, since $\kappa(-\infty) = \kappa(+\infty) = 0$,

$$\begin{aligned}\frac{1}{h}\int_{-\infty} x^r\, dx \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+\xi)\, d\xi &= \sum_{j=-\infty}^{\infty} \left\{ x_j^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j + \xi)\, d\xi \right\} - S_{m+1} \\ &= \bar{\mu}_r' - S_{m+1} \quad . \quad . \end{aligned} \tag{3.40}$$

The integral on the left of this expression is equal to

$$\frac{1}{h}\int_{-\infty}^{\infty}\int_{-\frac{h}{2}}^{\frac{h}{2}} x^r f(x+\xi)\, d\xi\, dx \quad . \tag{3.41}$$

provided that the multiple integral exists. If, in addition, it is absolutely convergent we may substitute x for $x + \xi$ and integrate with respect to ξ. We shall then have

$$\begin{aligned}\bar{\mu}_r' - S_{m+1} &= \frac{1}{h}\int_{-\infty}^{\infty}\int_{-\frac{h}{2}}^{\frac{h}{2}} (x-\xi)^r f(x)\, d\xi\, dx \\ &= \frac{1}{h}\int_{-\infty}^{\infty} f(x)\, dx \frac{\left(x+\frac{h}{2}\right)^{r+1} - \left(x - \frac{h}{2}\right)^{r+1}}{r+1} \\ &= \sum_{j=0}^{\left[\frac{r}{2}\right]} \left(\frac{h}{2}\right)^{2j} \binom{r}{2j} \frac{1}{2j+1} \mu_{r-2j}' \quad . \quad . \end{aligned} \tag{3.42}$$

where $\left[\frac{r}{2}\right]$ is the integral part of $\frac{r}{2}$.

Thus if S_{m+1} may be neglected, (3.42) gives the raw moments in terms of the actual moments. In practice we require the latter in terms of the former and it is easy to find from (3.42) the following expressions :—

$$
\left.
\begin{aligned}
\mu_1' &= \bar{\mu}_1' \\
\mu_2' &= \bar{\mu}_2' - \frac{1}{12}h^2 \\
\mu_3' &= \bar{\mu}_3' - \frac{1}{4}\bar{\mu}_1' h^2 \\
\mu_4' &= \bar{\mu}_4' - \frac{1}{2}\bar{\mu}_2' h^2 + \frac{7}{240}h^4 \\
\mu_5 &= \bar{\mu}_5' - \frac{5}{6}\bar{\mu}_3' h^2 + \frac{7}{48}\mu_1' h^4 \\
\mu_6' &= \bar{\mu}_6' - \frac{5}{4}\bar{\mu}_4' h^2 + \frac{7}{16}\bar{\mu}_2' h^4 - \frac{31}{1344}h^6
\end{aligned}
\right\}
\qquad (3.43)
$$

The general expression for these formulae is

$$\mu_r' = \sum_{j=0}^{r}\left\{\binom{r}{j}(2^{1-j} - 1)B_j h^j \bar{\mu}_{r-j}'\right\} \qquad (3.44)$$

where B_j is the Bernouilli number of order j. (Cf. Wold, 1934*a*.)

3.17. These are the corrections known as Sheppard's. It is important to realise the conditions under which they were obtained.

(*a*) It is assumed that $f(x)$ is bounded and tends monotonically to zero in the directions in which the range is infinite.

(*b*) It is assumed that the multiple integral (3.41) is absolutely convergent. This is equivalent to supposing that the absolute moment of order r exists. If $f(x)$ is finite in range and bounded, the multiple integral is certainly absolutely convergent. If the range is not finite, since $f(x)$ tends to zero monotonically in the direction or directions of infinite range,

$$\frac{1}{h}\int_{-\infty}^{\infty}\int_{-\frac{h}{2}}^{\frac{h}{2}} x^r f(x + \xi)\; d\xi\, dx$$

will converge or diverge with

$$\frac{1}{h}\int_{-\infty}^{\infty}\int_{-\frac{h}{2}}^{\frac{h}{2}} |\, x^r \,|\, f(x)\, d\xi\, dx$$

i.e. with

$$\int_{-\infty}^{\infty} |\, x^r \,|\, f(x)\, dx$$

which is the absolute moment of order r.

(*c*) It is assumed that $f(x)$ and its first m derivatives vanish at the terminal points of the range when the range is finite, or that

$$\lim\; x^r f^{(j)}(x) \to 0$$

for all j up to and including r when the range is infinite.

(*d*) It is assumed that S_{m+1} is negligible.

Now both $\dfrac{B_{2j}}{(2j)!}$ and $\dfrac{B_{2j}^{(1)}(\frac{1}{2})}{(2j)!}$ are less than $\dfrac{4}{6^{2j}}$ in magnitude * and hence S_{m+1} is of order $\dfrac{4nh^{m+1}}{6^m}$ multiplied by some value of $f^{(m)}(x)$ in the range. Thus if h is small, the range is finite and $f^{(m)}(x)$ is small; S_{m+1} will be small and may be neglected. In particular, if

$$2\left[\frac{r}{2}\right] < m \qquad . \quad . \quad . \quad . \quad . \quad . \quad (3.45)$$

the Sheppard corrections will give the moments accurate to order h^m, i.e. to the order of the terms applied in making the corrections.

3.18. The foregoing discussion is rigorous, but the corrections may be applied in practice with considerable confidence whenever there is high-order terminal contact.

Example 3.13

Consider the distribution

$$dF = \frac{1}{B(12,\ 6)}x^{11}(1-x)^5\,dx \qquad 0 \leqslant x \leqslant 1,$$

a case of the so-called Type I distribution. The exact frequencies for intervals of 0·1 may be obtained from the Tables of the Incomplete B-Function, and are as follows :—

Centre of Interval.	Frequency.
0·05	0·000,000,0
0·15	0,000,009,2
0·25	0·000,646,8
0·35	0·009,938,2
0·45	0·061,137,4
0·55	0·192,199,6
0·65	0·332,887,7
0·75	0·297,479,9
0·85	0·101,033,7
0·95	0·004,667,5
TOTAL	1·000,000,0

The raw moments about $x = 0$ are shown in the following table :—

Moment.	Raw.	Exact.	Corrected.
μ_1'	0·666,662,8	0·666,666,7	0·666,662,8
μ_2'	0·456,965,5	0·456,140,4	0·456,132,2
μ_3'	0·320,952,3	0·319,298,2	0·319,285,7
μ_4'	0·230,335,1	0·228,070,2	0·228,053,2
μ_5'	0·168,512,9	0·165,869,2	0·165,848,0
μ_6'	0·125,433,2	0·122,599,0	0·122,574,0

* For $B_{2j+1} = 0,\ j > 0$ and $\dfrac{B_{2j}}{2j!} = \dfrac{2(-1)^{j-1}}{(2\pi)^{2j}} \sum_{n=1}^{\infty} \left(\dfrac{1}{n^{2j}}\right.$

(cf. Milne Thomson, *loc. cit.*) $\dfrac{4(-1)^{j-1}}{(2\pi)^{2j}}$

and further $B_{2j}^{(1)}(\tfrac{1}{2}) \quad 1 - \dfrac{}{2^{2j-1}} \Big) B_{2j}.$

The exact values of the moments are calculable by evaluating integrals of the type $\int_0^1 x^{11+r}(1-x)^5\,dx$ and are shown in the third column. The final column shows the results obtained by applying equations (3.43), e.g.

$$\begin{aligned} \mu_2' &= \bar{\mu}_2' - h^2/12 \\ &= 0{\cdot}456{,}965{,}5 - 0{\cdot}000{,}833{,}3 \\ &= 0{\cdot}456{,}132{,}2. \end{aligned}$$

At the terminal $x = 1$, $f(x)$ and its derivatives up to the fourth vanish. At the other end, derivatives up to the tenth vanish. The function is bounded, of finite range, and the derivatives remain finite throughout the range. In virtue of (3.45) it is to be expected that corrected moments of third and lower orders will be accurate to the order of the terms in the corrections, i.e. μ_2' is accurate to order h^2 (0·001) and μ_3' to order h^3 (0·0001). Actually they are considerably more accurate than this. The corrected fourth moment is in error by a term of order 2×10^{-5}, and this is of the same magnitude as the correcting term $\frac{7}{240}h^4$ used in arriving at it. Similarly the corrected fifth moments are in error by a term of order 10^{-5}, of the same order as one of the correcting terms to the fifth moment, $\frac{7}{48}\mu_1' h^4$, and of the same order as or greater order than two correcting terms to the sixth moment, $\frac{7}{16}\mu_2' h^4$ and $-\frac{31}{1344}h^6$.

Thus the corrected moments are in all cases a substantial improvement on the raw moments; but in applying the corrections it is necessary to guard against being misled about the accuracy of the final result by the apparent precision of some of the small corrective terms.

Example 3.14

As an illustration of the way in which Sheppard's corrections break down when the condition for high-order contact is violated, an example is taken from a paper by Pairman and Pearson (1918). The following table shows the frequencies in a certain range of the normal distribution

$$dF = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$$

with intervals of width 0·5.

Interval centred at	Frequency.
1·5	0·655,91
2·0	0·278,34
2·5	0·092,45
3·0	0·024,02
3·5	0·004,89
4·0	0·000,78
4·5	0·000,10
5·0	0·000,01
TOTAL	1·056,50

The distribution has high-order contact at one end but not at the start of the curve, being in fact J-shaped and very abrupt at that point.

The following table shows the raw moments about the mean up to the fourth order,

the moments with Sheppard's corrections and the true moments calculated from the continuous normal distribution:—

Moment.	Raw.	Exact.	Corrected.
μ_2	0·158,524	0·172,222	0·137,691
μ_3	0·104,226	0·098,612	0·104,226
μ_4	0·149,090	0·156,405	0·131,097

It will be noted that in the two cases where the corrections are made they operate in the wrong direction. For the fourth moment they increase the difference between calculated and true values from about 4 per cent. to about 16 per cent. It is clear that, at least for the fairly coarse grouping of this example, Sheppard's corrections may fail completely.

3.19. Equations (3.43) were written in terms of moments about an arbitrary point. This point can, in particular, be the mean of the distribution, and accordingly we may drop the dashes and put μ_1' equal to zero in (3.43), to get the corrections appropriate for moments about the mean.

3.20. The discussion of the Sheppard corrections up to this point, and Examples 3.13 and 3.14, have supposed that the given frequencies were those of a distribution which was exactly specified by a continuous mathematical function. In practice this case very rarely occurs, the most common necessity for grouping corrections arising when moments are calculated from tables such as those of Chapter 1. For such tables it is not possible to state categorically that the corrections will result in an improvement; but there are usually strong presumptions to that effect. Consider, for example, the height data of Table 1.7 (Example 3.7). There can be no doubt that the histogram provided by this material can be graduated by a smooth curve and that such a curve will give better values of the moments than the histogram. Moreover, the tailing-off at the extremes of the distribution supports the assumption that the conditions for terminal contact are satisfied. It may therefore be confidently assumed that Sheppard's corrections as applied to the grouped data will give improved values for the exact values of the moments which would have been derived from the ungrouped data had they been available.

Average Corrections

3.21. There is a distinct type of problem which also leads to the Sheppard corrections. Suppose there is given a distribution of unknown range and the frequencies falling into specified intervals, one may ask what are the corrections to be applied to the raw moments so as to bring them *on the average* into closer relation with the real moments. In other words, supposing that the interval-mesh is located at random on the distribution, what are the average values of the raw moments?

Let X_j be a fixed set of values of x, j varying from $-\infty$ to ∞ by integral values. As x_j varies from $X_{j-\frac{h}{2}}$ to $X_{j+\frac{h}{2}}$, x_k varies from $X_{k-\frac{h}{2}}$ to $X_{k+\frac{h}{2}}$.

By definition

$$\bar{\mu}_r = \sum_{j=-\infty}^{\infty} \left\{ x_j^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j + \xi)\, d\xi \right\}$$

Denoting by $E(\bar{\mu}'_r)$ the average as x_j varies from $X_{j-\frac{h}{2}}$ to $X_{j+\frac{h}{2}}$, we have

$$E(\bar{\mu}'_r) = \frac{1}{h}\int_{x_j-\frac{h}{2}}^{x_j+\frac{h}{2}} \sum\left\{x_j^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j+\xi)\,d\xi\right\} dx_j$$

$$= \frac{1}{h}\sum_{j=-\infty}^{\infty}\int_{x_j-\frac{h}{2}}^{x_j+\frac{h}{2}} x_j^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j+\xi)\,d\xi\,dx_j$$

$$= \frac{1}{h}\int_{-\infty}^{\infty} x^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+\xi)\,d\xi\,dx \quad . \quad . \quad . \quad (3.46)$$

which is the same as equation (3.40) with the omission of S_{m+1} and the substitution of $E(\bar{\mu}'_r)$ for $\bar{\mu}'_r$. Thus the Sheppard corrections apply for the average group-moments whatever the nature of the terminal contact.

They cannot, however, be applied indiscriminately on that ground. In place of the conditions about terminal contact, which ensure the applicability of Sheppard's corrections to any particular distribution, there is the condition that the grouping intervals are located at random on the range, which implies that although the corrections may be wrong in any given instance, the average effect in a large number of cases will be correct. In actual fact the condition about the random location of grouping does not operate very frequently for J- and U-shaped distributions, where the Sheppard corrections would not ordinarily apply ; for instance, in a distribution of incomes or deaths at given ages it is almost inevitable to begin the grouping at zero.

3.22. It is also illegitimate to drop the dashes in order to obtain corrections for moments about the mean. If the mean of the grouped distribution is denoted by y, the average value of the rth moment about the mean is given by

$$E(\bar{\mu}_r) = \frac{1}{h}\int_{-\infty}^{\infty}(x-y)^r\int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+\xi)\,d\xi\,dx,$$

where y is a function of x and the transformation of the integral which has been used earlier in this chapter is no longer legitimate. Explicit expressions for average corrections to moments about the mean have not yet been obtained. From a consideration of some particular distributions, however, Kendall (1938) concluded that for all ordinary purposes it is sufficient to use equations (3.43) as if the mean were a fixed point.

3.23. The Sheppard corrections have also been considered from a slightly different point of view (Fisher, 1922). As the centres of the intervals move along the variate axis, the raw moments vary according to the different groupings which result ; and this variation is evidently periodic of period h. We may thus write

$$\bar{\mu}'_r = \sum_{j=-\infty}^{\infty} \zeta^r \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} f(x)\,dx,$$

where
$$\zeta = \left(j + \frac{\theta}{2\pi}\right)h$$

and may put this equal to

$$A_0 + A_1 \sin\theta + A_2 \sin 2\theta + \ldots$$
$$+ B_1 \cos\theta + B_2 \cos 2\theta + \ldots$$

Then, multiplying by $\sin s\theta$ or $\cos s\theta$ and integrating from 0 to 2π, we have

$$A_s = \frac{1}{\pi}\sum_{j=-\infty}^{\infty}\int_0^{2\pi} \sin s\theta\, d\theta \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} \zeta^r f(x)\, dx$$

$$B_s = \frac{1}{\pi}\sum_{j=-\infty}^{\infty}\int_0^{2\pi} \cos s\theta\, d\theta \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} \zeta^r f(x)\, dx$$

and in particular

$$A_0 = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\int_0^{2\pi} d\theta \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} \zeta^r f(x)\, dx.$$

Since $d\theta = \frac{2\pi}{h} d\zeta$, we have

$$\begin{aligned}
A_0 &= \frac{1}{h}\sum_{j=-\infty}^{\infty}\int_{jh}^{(j+1)h} d\zeta \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} \zeta^r f(x)\, dx \\
&= \frac{1}{h}\int_{-\infty}^{\infty} d\zeta \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} \zeta^r f(x)\, dx \\
&= \frac{1}{h}\int_{-\infty}^{\infty} f(x)\, dx \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} \zeta^r\, d\zeta \\
&= \frac{1}{h}\int_{-\infty}^{\infty} x^r \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+\xi)\, d\xi\, dx
\end{aligned}$$

which is the same as (3.41) and (3.46), and thus leads to the Sheppard corrections.

For the periodic terms we have

$$\begin{aligned}
A_s &= \frac{2}{h}\int_{-\infty}^{\infty} \sin\frac{2\pi s\zeta}{h}\, d\zeta \int_{\zeta-\frac{h}{2}}^{\zeta+\frac{h}{2}} \zeta^r f(x)\, dx \\
&= \frac{2}{h}\int_{-\infty}^{\infty} f(x)\, dx \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} \zeta^r \sin\frac{2\pi s\zeta}{h}\, d\zeta.
\end{aligned}$$

For some mathematically specified distributions we are able to consider the magnitude

of these periodic terms. For instance, for the normal curve referred to the true mean we have, since

$$\frac{2}{h}\int_{x-\frac{h}{2}}^{x+\frac{h}{2}} \zeta \sin\frac{2\pi s\zeta}{h}\,d\zeta = -\frac{h}{\pi s}\cos\frac{2\pi sx}{h}\cos\pi s$$

$$_1A_s = (-1)^{s+1}\frac{h}{\pi s}\int_{-\infty}^{\infty}\cos\frac{2\pi sx}{h}\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}\,dx$$

$$= (-1)^{s+1}\frac{h}{\pi s}e^{\frac{2s^2\sigma^2\pi^2}{h^2}}$$

$$_1B_s = 0,$$

where $_1A_s$ and $_1B_s$ refer to the coefficients for the corrections to the mean. The grouping error of the mean is thus

$$\frac{h}{\pi}\left(e^{-\frac{2\sigma^2\pi^2}{h^2}}\sin\theta - \tfrac{1}{2}e^{-\frac{8\sigma^2\pi^2}{h^2}}\sin 2\theta + \quad . \quad . \quad . \text{ etc.}\right).$$

For a grouping in which $\sigma = h$ (a very coarse grouping) this is, approximately, $-\frac{\sigma}{\pi}e^{-2\pi^2}\sin\theta$ and thus cannot be greater than $\frac{\sigma}{\pi}e^{-2\pi^2}$.

3.24. Average corrections may also be applied to discrete data which have been grouped in wider intervals but are different from those of the continuous case. Cf. Exercise 3.13 and C. C. Craig (1936).

Sheppard's Corrections to Factorial Moments

3.25. It has been shown by Wold (1934*a*) that for factorial moments the Sheppard corrections are as follows :—

$$\begin{aligned}
\mu'_{[1]} &= \bar{\mu}'_{[1]} \\
\mu'_{[2]} &= \bar{\mu}'_{[2]} - \frac{h^2}{12} \\
\mu'_{[3]} &= \bar{\mu}'_{[3]} - \frac{h^2}{4}\bar{\mu}'_{[1]} + \frac{h^3}{4} \\
\mu'_{[4]} &= \bar{\mu}'_{[4]} - \frac{h^2}{2}\bar{\mu}'_{[2]} + h^3\bar{\mu}'_{[1]} - \frac{71}{80}h^4 \\
\mu'_{[5]} &= \bar{\mu}'_{[5]} - \frac{5}{6}\bar{\mu}'_{[3]}h^2 + \frac{5}{2}\bar{\mu}'_{[2]}h^3 - \frac{71}{16}\bar{\mu}'_{[1]}h^4 + \frac{31}{8}h^5 \\
\mu'_{[6]} &= \bar{\mu}'_{[6]} - \frac{5}{4}\bar{\mu}'_{[4]}h^2 + 5\bar{\mu}'_{[3]}h^3 - \frac{213}{16}\bar{\mu}'_{[2]}h^4 \\
&\quad + \frac{93}{4}\bar{\mu}'_{[1]}h^5 - \frac{9129}{448}h^6
\end{aligned} \qquad (3.47)$$

and in general are given by

$$\mu'_{[r]} = \sum_{j=0}^{r}\binom{r}{j}B_j^{(j+2)}\left(\frac{3}{2}\right)h^j\bar{\mu}'_{[r-j]}, \qquad (3.48)$$

where the Bernouilli polynomial $B_j^{(j+2)}(\frac{3}{2})$ is equal to

$$(-1)^{j+1}\frac{(2j)!}{2^{2j}(j+1)!} \quad + \tfrac{1}{5} + \dots + \frac{1}{2j-1}\Bigg), \qquad j > 1$$

and
$$B_0^{(2)}(\tfrac{3}{2}) = 1, \quad B_1^{(3)}(\tfrac{3}{2}) = 0.$$

Sheppard's Corrections for Cumulants

3.26. As in section **3.16**, and under the same conditions, we have, writing θ for it,

$$\sum_{j=-\infty}\Big\{e^{\theta x_j}\int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_j+\xi)\,d\xi\Big\} = \frac{1}{h}\int_{-\infty}^{\infty} e^{\theta x}\,dx\int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+\xi)\,d\xi$$
$$= \frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} e^{-\theta\xi}\,d\xi\int_{-\infty}^{\infty} e^{\theta x}f(x)\,dx$$
$$= \frac{\sinh\frac{\theta h}{2}}{\theta h}\int_{-\infty}^{\infty} e^{\theta x}f(x)\,dx. \qquad (3.49)$$

The expression on the left gives the characteristic function for the grouped data, and the integral on the right the true characteristic function. Taking logarithms of both sides and noting * that

$$\log\frac{\sinh\frac{\theta h}{2}}{\frac{\theta h}{2}} = \sum_{r=2}^{\infty}\frac{B_r(\theta h)^r}{r!r}$$

we have, for the coefficient in $\frac{}{r!}$, $\kappa_r = \bar{\kappa}_r \quad B_r h^r, \; r > 1$. $\qquad$ (3.50)

an attractively simple result for the Sheppard corrections to cumulants. Since all B's of odd order are zero except B_1 and the first cumulant is equal to the mean, no cumulant of odd order needs any correction. For the others we have

$$\begin{aligned}\kappa_2 &= \bar{\kappa}_2 - \frac{h^2}{12}\\ \kappa_4 &= \bar{\kappa}_4 + \frac{h^4}{120}\\ \kappa_6 &= \kappa_6 - \frac{h^6}{252}\end{aligned} \qquad (3.51)$$

* By definition
$$\frac{\theta}{\quad -1} = \sum_{r=0}^{\infty}\frac{B_r\theta^r}{r!}$$

and hence
$$\frac{1}{\quad -1} - \frac{1}{\theta} + \frac{1}{2} = \sum^{\infty}\frac{B_r\theta^{r-1}}{r!};$$

integrating from 0 to θ we have the above result.

Grouping Corrections when the Distribution is Abrupt

3.27. Various writers have considered the corrections to be applied when one or both terminals of the distribution do not obey the Sheppard conditions for terminal contact. References are given at the end of this chapter.

Multivariate Moments and Cumulants

3.28. The foregoing results in this chapter may be readily generalised to the multivariate case. To save complicating the algebraic expressions, we shall deal with two variates x_1 and x_2; but the reader will have little difficulty in carrying out any generalisations for more variates.

The bivariate moment μ'_{rs} about an origin a_1 for x_1 and a_2 for x_2 is defined by

$$\mu'_{rs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 - a_1)^r(x_2 - a_2)^s\,dF \tag{3.52}$$

If one of r, s is zero the moment becomes the ordinary univariate moment of the row or column-border distribution of the bivariate population. In the contrary case we meet a new type of moment—the product-moment. The first product-moment μ'_{11} is of particular importance in the theory of correlation. The first product-moment about the variate means, μ_{11}, is known as the Covariance.

As in the univariate case, bivariate moments about certain points can be expressed in terms of those about other points. If the x_1 origin is transferred from a_1 to b_1 where $c_1 = b_1 - a_1$, and the x_2 origin from a_2 to b_2, where $c_2 = b_2 - a_2$, we have

$$\mu'_{rs}(a_1\,a_2) = (\mu' + c_1)^r(\mu' + c_2)^s \tag{3.53}$$

where the product $\mu'^j\mu'^k$ on the right is to be replaced by $\mu'_{jk}(b_1\,b_2)$. This corresponds to the symbolic equation

$$\mu^r(a) = \{\mu'(b) + c\}^r$$

for the univariate case.

Methods of calculating the product-moments for numerically specified distributions will be considered in Chapter 14. The determination of bivariate moments for a mathematically specified population is a matter of evaluating double sums or double integrals, and no new statistical points call for comment.

Example 3.15

The bivariate distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}dx_1\,dx_2 \qquad -\infty \leqslant x_1, x_2 \leqslant \infty.$$

Let us evaluate:—

$$M(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1 t_1 + x_2 t_2}\,dF.$$

Making the substitution,

$$\xi = x_1 - \sigma_1^2 t_1 - \rho\sigma_1\sigma_2 t_2$$
$$\eta = x_2 - \rho\sigma_1\sigma_2 t_1 - \sigma_2^2 t_2$$

we find

$$M(t_1, t_2) = \exp\{\tfrac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\sigma_1\sigma_2\rho + t_2^2\sigma_2^2)\} \times$$

$$\frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\xi^2}{\sigma_1^2} - \frac{2\rho\xi\eta}{\sigma_1\sigma_2} + \frac{\eta^2}{\sigma_2^2}\right)\right\}d\xi\, d\eta$$

$$= \exp\{\tfrac{1}{2}(t_1^2\sigma_1^2 + 2\rho\sigma_1\sigma_2t_1t_2 + t_2^2\sigma_2^2)\}.$$

Now μ_{rs} is the coefficient of $\dfrac{t_1^r}{r!}\dfrac{t_2^s}{s!}$ in $M(t_1, t_2)$ and thus we find, for instance,

$$\begin{aligned}
&\mu_{20} = \sigma_1^2, \quad \mu_{11} = \rho\sigma_1\sigma_2, \quad \mu_{02} = \sigma_2^2\\
&\mu_{30} = \mu_{21} = \mu_{12} = \mu_{03} = 0\\
&\mu_{40} = 3\sigma_1^4, \quad \mu_{31} = 3\rho\sigma_1^3\sigma_2, \quad \mu_{22} = (1+2\rho^2)\sigma_1^2\sigma_2^2,\\
&\mu_{13} = 3\rho\sigma_1\sigma_2^3, \quad \mu_{04} = 3\sigma_2^4.
\end{aligned}$$

3.29. The bivariate analogue of equation (3.22) may be written

$$\exp\left\{\frac{\kappa_{10}t_1}{1!0!} + \frac{\kappa_{01}t_2}{0!1!} + \ldots \frac{\kappa_{rs}}{r!s!}t_1^rt_2^s + \ldots\right.$$

$$= 1 + \frac{\mu'_{10}t_1}{1!0!} + \frac{\mu'_{01}t_2}{0!1!} + \ldots + \frac{\mu'_{rs}}{r!s!}t_1^rt_2^s + \qquad (3.54)$$

or symbolically,

$$\exp\left\{\Sigma\frac{1}{p!}\kappa(t_1+t_2)^p\right\} = \Sigma\frac{1}{p!}\mu(t_1+t_2)^p \qquad (3.55)$$

where

$$\kappa(t_1+t_2)^p = p!\left\{\frac{\kappa_{p0}t_1^p}{p!0!} + \frac{\kappa_{p-1,\,1}\,t_1^{p-1}t_2}{(p-1)!1!} + \ldots\right. \qquad (3.56)$$

In terms of characteristic functions we may define

$$\phi(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ix_1t_1+ix_2t_2}\, dF \qquad (3.57)$$

and, as before, write

$$\phi(t_1, t_2) = \sum_{r,s=0}^{\infty}\mu'_{rs}\frac{(it_1)^r}{r!}\frac{(it_2)^s}{s!}$$

$$= \exp\left\{\sum_{r,s=1}^{\infty}\kappa_{rs}\frac{(it_1)^r}{r!}\frac{(it_2)}{s!}\right. \qquad (3.58)$$

subject to conditions of existence.

From these equations the bivariate moments can be expressed in terms of bivariate cumulants and vice-versa. It is also possible to derive bivariate equations from the univariate equations by symbolic processes (cf. Kendall, 1941).

3.30. Wold (1934*b*) has given the following expressions for Sheppard corrections to bivariate moments and cumulants, the variates being grouped in intervals h_1, h_2.

$$\mu'_{rs} = \sum_{j=0}^{r}\sum_{k=0}^{s}h_1^jh_2^k\binom{r}{j}\binom{s}{k}(2^{1-j}-1)(2^{1-k}-1)B_jB_k\,\bar{\mu}'_{r-j,s-k} \qquad (3.59)$$

In particular

$$\left.\begin{aligned}
&\mu'_{20} = \bar{\mu}'_{20} - \tfrac{1}{12}h_1^2, \quad \mu'_{11} = \bar{\mu}'_{11}, \quad \mu'_{02} = \bar{\mu}'_{02} - \tfrac{1}{12}h_2^2\,;\\
&\mu'_{30} = \bar{\mu}'_{30} - \mu'_{10}\frac{h_1^2}{4}, \ \mu'_{21} = \bar{\mu}'_{21} - \bar{\mu}'_{01}\frac{h_1^2}{12}, \text{ and two symmetrical equations}\\
&\mu'_{40} = \bar{\mu}'_{40} - \bar{\mu}'_{20}\frac{h_1^2}{2} + \frac{7}{240}h_1^2, \ \mu'_{31} = \bar{\mu}'_{31} - \bar{\mu}'_{11}\frac{h_1^2}{4}, \text{ and two symmetrical equations}\\
&\mu'_{22} = \bar{\mu}'_{22} \quad \bar{\mu}'_{20}\frac{h_2^2}{12} - \bar{\mu}'_{02}\frac{h_1^2}{12} + \frac{h_1^2h_2^2}{144}
\end{aligned}\right\} \quad . \quad (3.60)$$

For cumulants we have

$$\left.\begin{aligned}
\kappa_{rs} &= \bar{\kappa}_{rs}, & r, s &> 0\\
\kappa_{r0} &= \bar{\kappa}_{r0} - \frac{B_r h_1^r}{r} & r &\geqslant 2\\
\kappa_{0s} &= \bar{\kappa}_{0s} - \frac{B_s h_2^s}{s} & s &\geqslant 2
\end{aligned}\right\} \quad . \quad (3.61)$$

Measures of Skewness

3.31. We have considered measures of location and dispersion in Chapter 2. With the aid of the moments we can now proceed to consider measures of other qualities of the population, and in particular its departure from symmetry.

In a symmetrical population, mean, median and mode coincide. It is thus natural to take the deviation mean to mode or mean to median as measuring the skewness of the distribution. K. Pearson proposed the measure

$$Sk = \text{Mean} - \text{mode}$$

which is subject to the inconvenience of determining the mode. For a wide class of frequency-distributions known as Pearson's (cf. Chapter 6), this measure may, however, be expressed exactly in terms of the first four moments. We define

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad . \quad (3.62)$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad (3.63)$$

Then it may be shown that for Pearson curves

$$Sk = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)} \quad (3.64)$$

and this equation may be taken as defining a measure of skewness applicable to all distributions whose moments up to and including the fourth exist.

The coefficient β_1 itself is also a measure of skewness. Clearly if the distribution is symmetrical it vanishes since μ_3 vanishes, and the size of μ_3 relative to $\mu_2^{\frac{3}{2}}$ (or $\sqrt{\beta_1}$) will indicate the extent of the departure from symmetry.

Generally we may define

$$\beta_{2n+1} = \frac{\mu_3\mu_{2n+3}}{\mu_2^{n+3}} \qquad \qquad (3.65)$$
$$\beta_{2n} = \frac{\mu_{2n+2}}{\mu_2^{n+1}}$$

quantities which are not in general use but will be found to occur occasionally in statistical literature.

More convenient quantities than β_1 and β_2 for certain purposes are

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{\kappa_3}{\kappa_2^{\frac{3}{2}}} \qquad . \quad (3.66)$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\kappa_4}{\kappa_2^2} \qquad . \quad . \quad . \quad . \quad . \quad (3.67)$$

If the distribution is expressed in standard measure, γ_1 and γ_2 are its third and fourth cumulants.

Kurtosis

3.32. In the so-called "normal" distribution

$$dF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} dx, \qquad -\infty \leqslant x \leqslant \infty$$

β_2 attains the value 3 and γ_2 is zero. Curves for which $\gamma_2 = 0$ are called Mesokurtic. Those for which $\gamma_2 > 0$ are called Leptokurtic and will, relative to the normal curve, be sharply peaked. Those for which $\gamma_2 < 0$ are called Platykurtic and will be flat-topped.

Example 3.16

For the distribution of Australian marriages considered in Example 3.1 we found, for the raw moments about the mean in units of three years,

$$\bar{\mu}_2 = 7{\cdot}056{,}977, \quad \bar{\mu}_3 = 36{\cdot}151{,}595, \quad \bar{\mu}_4 = 408{\cdot}738{,}210.$$

With Sheppard's corrections these become

$$\mu_2 = 6{\cdot}973{,}644, \quad \mu_3 = 36{\cdot}151{,}595, \quad \mu_4 = 405{\cdot}238{,}888.$$

From these values we find

$$\beta_1 = 3{\cdot}854, \quad \gamma_1 = 1{\cdot}963$$
$$\beta_2 = 8{\cdot}333, \quad \gamma_2 = 5{\cdot}333$$

indicating considerable skewness and leptokurtosis.

Example 3.17

From the formulae for the moments of the binomial distribution considered in Example 3.2 we find

$$\gamma_1 = \frac{q-p}{\sqrt{npq}}$$
$$\gamma_2 = \frac{1-6pq}{npq}$$

so that, as $n \to \infty$, γ_1 and $\gamma_2 \to 0$. This is in accordance with a result we shall prove later, that the binomial tends to the normal form as n tends to infinity.

Moments as Characteristics of a Distribution

3.33. The use of moments and cumulants in determining the nature of a frequency-distribution will be abundantly illustrated in later chapters, but some general remarks may be made at this stage.

It has been noted that the characteristic function determines the moments when they exist, and it will be proved in Chapter 4 that the characteristic function also determines the distribution function. It does not, however, follow that the moments completely determine the distribution, even when moments of all orders exist. Only under certain conditions will a set of moments determine a distribution uniquely, but, fortunately for statisticians, those conditions are obeyed by all the distributions arising in statistical practice. For all ordinary purposes, therefore, a knowledge of the moments, when they all exist, is equivalent to a knowledge of the distribution function: equivalent, that is, in the sense that it should be possible *theoretically* to exhibit all the properties of the distribution in terms of the moments.

3.34. In particular we expect that if two distributions have a certain number of moments in common they will bear some resemblance to each other. If, say, moments up to those of order n are identical we know that as n tends to infinity the distributions approach identity, and consequently we expect that by identifying the lower moments of two distributions we bring them to approximate equality. Some mathematical support for this so-called Principle of Moments may be derived from the following approach:

It is known that a function which is continuous in a finite range a to b can be represented in that range by a uniformly convergent series of polynomials in x, say $\sum_{n=0}^{\infty} P_n(x)$ where $P_n(x)$ is of degree n. Suppose we wish to represent such a function approximately by the *finite* series of powers $\sum_{n=0}^{n} a_n x^n$. The coefficients a_n may be determined by the principle of least squares, i.e. so as to make

$$\int_a^b (f - \Sigma a_n x^n)^2 \, dx \qquad . \quad . \quad . \quad . \quad . \quad . \quad (3.68)$$

a minimum. Differentiating by a_j we have

$$2\int_a^b (f - \Sigma a_n x^n) a_j x^j \, dx = 0$$

or

$$\int_a^b f x^j \, dx = \mu_j' = \int_a^b \Sigma a_n x^{n+j} \, dx. \qquad . \quad . \quad . \quad . \quad . \quad (3.69)$$

If now two distributions have moments up to order n equal they must have the same least-squares approximation, for the coefficients a_n are determined by the moments in virtue of (3.69). Furthermore, if in the range the distribution f_1 differs from $\Sigma a_n x^n$ by ε_1 and f_2 by ε_2, then f_1 differs from f_2 by not more than $\varepsilon_1 + \varepsilon_2$.

A similar line of approach may be adopted when the range is infinite, the distributions in such cases being, under certain general conditions, capable of representation by a series of terms such as $e^{-x^2} P_n(x)$. (Cf. Chapter 6.) The same conclusion is reached.

Thus distributions which have a finite number of the lower moments in common will, in a sense, be approximations one to another. We shall encounter many cases where, although we cannot determine a distribution function explicitly, we may ascertain its

moments at least up to some order; and hence we shall be able to approximate to the distribution by finding another distribution of known form which has the same lower moments. In practice, approximations of this kind often turn out to be remarkably good, even when only the first three or four moments are equated.

Mean Values

3.35. To conclude this chapter we may note that the moments are particular cases of a general class of functions known as Mean Values. If we have a function $\psi(x)$ defined in the range of a distribution, then

$$\psi(x)\, dF, \tag{3.70}$$

if it exists, is called the mean value of $\psi(x)$ for that distribution; it is sometimes written as $E\{\psi(x)\}$, a notation we shall often find useful. The moment of order r is thus the mean value of x^r and the characteristic function is the mean value of e^{itx}. The letter E in this connection is the first of the word "expectation," and mean values as we have defined them are sometimes known as "expected" values, particularly in the theory of probability. The objection to this practice is that only rarely is it to be expected that we shall meet with the "expected" value in sampling.

3.36. Two important properties of mean values are to be noted. In the first place, if we have two functions $\psi_1(x)$ and $\psi_2(x)$,

$$\int \psi_1\, dF + \int \psi_2\, dF = \int (\psi_1 + \psi_2)\, dF$$

and thus

$$E(\psi_1 + \psi_2) = E(\psi_1) + E(\psi_2), \tag{3.71}$$

i.e. the mean value of a sum is the sum of the mean values.

Secondly, if we have two *independent* variates x_1 and x_2 distributed with functions F_1, F_2; and if ψ_1 is a function of x_1 and ψ_2 of x_2, then

$$\int\int \psi_1\psi_2\, dF_1\, dF_2 = \int \psi_1\, dF_1 \int \psi_2\, dF_2$$

or,

$$E(\psi_1\psi_2) = E(\psi_1)E(\psi_2) \tag{3.72}$$

so that the mean value of the product is the product of the mean values. This is in general only true if the variates are independent, whereas (3.71) is subject to no such restriction.

NOTES AND REFERENCES

In most of the literature what have here been called "cumulants" are referred to as semi-invariants or seminvariants. They were introduced by Thiele (1889), who, however, failed to draw a clear distinction between the parameters of a population and estimates of those parameters from a sample, with the result that for some years there was a confusion between semi-invariant parameters and semi-invariant statistics. (This is in no way to be interpreted as a criticism of Thiele, who could hardly have been expected to write fifty years ahead of his time.) Some recent work by Dressel (1940) has shown the desirability of reserving the name "seminvariant" for the more general class of parameters which

are, except for powers of κ_2, invariant under transformations of the origin. Dressel points out the analogy between such parameters and the functions of the coefficients of the binary form

$$a_0x^n + \binom{n}{1}a_1x^{n-1}y + . \qquad \binom{n}{r}a_rx^{n-r}y^r + . . . a_ny^n,$$

which are invariant under transformations of type

$$\xi = lx + m, \quad y = \eta.$$

The word "seminvariant" has been in use for many years in the theory of algebraic invariants to denote such functions. The word "cumulant" is due to Fisher and Wishart.

A comprehensive account of the mathematical relations between moments, factorial moments and cumulants is given by Frisch (1926).

There is an extensive literature on corrections for grouping. Kendall (1938) gave a bibliography which appears to be complete except for the omission of a paper by Fisher (1922) and one by Elderton (1938*b*). For corrections in the case when the Sheppard conditions are violated, see Pairman and Pearson (1918), Sandon (1924), Martin (1934), Pearse (1928) and Elderton (1938*a*). For Sheppard's corrections for a discrete variable (which appear to be due to H. C. Carver) see Craig (1936); and for the corrections in the multivariate case see Wold (1934*b*).

References to the problem of moments (i.e. the conditions under which a set of constants can form the moments of a distribution) are given at the end of Chapter 4. As to the mathematical basis of the principle of moments, see Merzrath (1933) and Romanovsky (1936).

Craig, C. C. (1936), "Sheppard's corrections for a discrete variable," *Ann. Math. Statist.* **7**, 55.

Dressel, P. L. (1940), "Statistical seminvariants and their estimates with particular emphasis on their relation to algebraic invariants," *Ann. Math. Statist.*, **11**, 33.

Elderton, Sir W. P. (1938*a*), *Frequency Curves and Correlation*, Cambridge University Press.

—— (1938*b*), "Correzioni dei momenti quando la curva è simmetrica," *Giorn. dell' Ist. Ital. Att.*, **16**, 145.

Fisher, R. A. (1922), "On the mathematical foundations of theoretical statistics," *Phil. Trans.*, A, **222**, 309.

Frisch, R. (1926), "Sur les semi-invariants et moments employés dans l'étude des distributions statistiques," Oslo; *Skrifter af det Norske Videnskaps-Akademie, II, Hist.-Filos. Klasse, No. 3.*

Kendall, M. G. (1938), "The conditions under which Sheppard's corrections are valid," *J. Roy. Statist. Soc.*, **101**, 592.

—— (1941), "The derivation of multivariate sampling formulae from univariate formulae by symbolic operation," *Ann. Eugen. Lond.*, **10**, 392.

Martin, E. S. (1934), "On the correction for moment coefficients of frequency distributions when the start of the frequency is one of the characteristics to be determined," *Biometrika*, **26**, 12.

Merzrath, E. (1933), "Anpassung von Flächen an zweidimensionale Kollektivgegenstände und ihre Auswertung für die Korrelationstheorie," *Metron*, **11**, No. 2, 103.

Pairman, E., and Pearson, K. (1918), "On corrections for the moment-coefficients of limited range frequency-distributions when there are finite or infinite ordinates and any slopes at the terminals of the range," *Biometrika*, **12**, 231.

Pearse, G. E. (1928), "On corrections for the moment-coefficients of frequency-distributions when there are infinite ordinates at one or both terminals of the range," *Biometrika*, **20A**, 314.

Romanovsky, V. (1936), "Note on the method of moments," *Biometrika*, **28**, 188.

Sandon, F. (1924), "Note on the simplification of the calculation of the abruptness coefficients to correct crude moments," *Biometrika*, **16**, 193.

Shohat, J. (1929), "Inequalities for moments of frequency functions and for various statistical constants," *Biometrika*, **21**, 361.

Thiele, T. N. (1889), *Theory of Observations* (reprinted in English in *Ann. Math. Statist.*, 1931, **2**, 165).

Wold, H (1934*a*). "Sulla correzione di Sheppard," *Giorn. dell' Ist. Ital. Att.*, **4**, 304.

——— (1934*b*), "Sheppard's correction formulae in several variables," *Skand. Aktuar.*, **17**, 248.

EXERCISES

3.1. Show that the rth moment about the origin of the distribution

$$dF = kx^{-p}e^{-\gamma/x}\,dx \qquad 0 \leqslant x \leqslant \infty, \quad \gamma > 0$$

is

$$\mu'_r = \frac{\gamma^r \Gamma(p-r-1)}{\Gamma(p-1)}$$

if $r < p - 1$, and does not exist in the contrary case.

3.2. In the distribution

$$dF = k\left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-\nu \tan^{-1}\frac{x}{a}} \qquad -\infty \leqslant x \leqslant \infty$$

show that, about the origin,

$$\mu'_r = ka^{r+1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{2m-r-2}\theta \sin^r\theta\, e^{-\nu\theta}\,d\theta$$

and hence that

$$\mu'_r = \frac{a}{2m-r-1}\{(r-1)a\mu'_{r-2} - \nu\mu'_{r-1}\}.$$

3.3. Show that the discontinuous distribution whose frequencies corresponding to the values 0, 1, . . . j . . . are

$$e^{-m}\left(1, \frac{m}{1!}, \ldots \frac{m^j}{j!} \ldots\right)$$

has, for the moments about the mean,

$$\mu_2 = m,\ \mu_3 = m,\ \mu_4 = m(1+3m),\ \mu_5 = m(1+10m),\ \mu_6 = m(1+25m+15m^2).$$

3.4. Show that for the distribution whose frequencies for variate-values 0, 1, . . . j, . . . are the successive terms in $(\frac{1}{2}+\frac{1}{2})^n$, i.e. $(\frac{1}{2})^n\left[1, \binom{n}{1}, \binom{n}{2} \ldots\right]$, all cumulants of odd order except the first vanish.

3.5. Show generally that the cumulants of odd order vanish for any symmetrical distribution, except the first.

3.6. Show that e^{itx} may be expanded in an infinite series, valid in $-\infty \leqslant x \leqslant \infty$,

$$1 + (e^{it}-1)\, x^{[1]} + (e^{it}-1)^2 \frac{x^{[2]}}{2!} + \ldots (e^{it}-1)^r \frac{x^{[r]}}{r!} + \ldots$$

the factorials being taken with unit interval; and hence that

$$\mu'_{[r]} = [d^r \phi(t)]_{t=0}$$

where

$$d = \frac{d}{d(e^{it})}.$$

Hence show that, for the binomial $(q+p)^n$ about the origin,

$$\mu'_{[r]} = n^{[r]} p^r.$$

3.7. Show that the distribution whose frequency at the variate-value $\pm 2r$ (r integral) is

$$e^{-2a}\left\{\frac{a^{2r}}{0!(2r)!} + \frac{a^{2r+2}}{1!(2r+1)!} + \frac{a^{2r+4}}{2!(2r+2)!} + \ldots\right.$$

and at $\pm(2r+1)$ is

$$e^{-2a}\left.\frac{a^{2r+1}}{(2r+1)!} + \frac{a^{2r+3}}{1!(2r+3)!} + \frac{a^{2r+5}}{2!(2r+5)!} + \ldots\right\}$$

has odd-order cumulants equal to zero and even-order cumulants equal to $2a$.

3.8. Show that for the distribution

$$dF = \frac{1}{\sigma} e^{-\frac{x}{\sigma}} dx, \qquad 0 \leqslant x \leqslant \infty$$

$$\kappa_r = \sigma^r (r-1)!$$

3.9. Show that

$$\mu'_r = \sum_{j=1}^{r} \binom{r-1}{j-1} \mu'_{r-j} \kappa_j$$

and hence that

$$\kappa_r = (-1)^{r-1} \begin{array}{llllll} \mu_1 & & & 0 & & 0 \\ \mu'_2 & \binom{1}{0}\mu'_1 & 1 & 0 & & 0 \\ \mu'_3 & \binom{2}{0}\mu'_2 & \binom{2}{1}\mu'_1 & & & \\ & \cdot & \cdot & & & \\ \mu'_r & \binom{r-1}{0}\mu'_{r-1} & \binom{r-1}{1}\mu'_{r-2} & & & \binom{r-1}{r-2}\mu'_1 \end{array}$$

3.10. Show that for the distribution

$$dF = dx, \qquad 0 \leqslant x \leqslant 1$$

grouped into an integral number of intervals of equal width h, the corrections to the second and fourth moments about the mean are

$$\mu_2 = \bar{\mu}_2 + \frac{h^2}{12}$$

$$\mu_4 = \bar{\mu}_4 + \bar{\mu}_2\frac{h^2}{12} + \frac{h^4}{80}.$$

(Cf. Elderton, 1938*b*. Note that the first is exactly, and the second approximately, the Sheppard correction *with sign reversed.*)

3.11. If ∂_p stands for the operator such that

$$\begin{aligned} \partial_p\mu'_r &= r^{[p]}\mu'_{r-p}, & r \geqslant p \\ &= 0 & r < p \end{aligned}$$

and ∂_p is distributive when applied to products, e.g.

$$\partial_p(AB) = B(\partial_p A) + A(\partial_p B),$$

show that ∂_p annihilates every cumulant (considered as a function of the moments) except κ_p, and that

$$\partial_p\kappa_p = p!$$

(Cf. Kendall, 1941.)

3.12. If $f(x)$ is an odd function of x of period $\frac{1}{2}$, show that

$$\int_0^\infty x^r x^{-\log x} f(\log x)\, dx = 0$$

for all integral values of r. Hence show that the distributions

$$dF = x^{-\log x}\{1 - \lambda \sin(2\pi \log x)\}dx \qquad \begin{matrix} 0 \leqslant x \leqslant \infty \\ 0 \leqslant \lambda \leqslant 1 \end{matrix}$$

have the same moments whatever the value of λ. (Stieltjes. See refs. to Chapter 4.)

3.13. Show that if the frequencies of a discontinuous distribution are distributed at equal intervals $\frac{h}{m}$, m in each grouping interval h, the average grouping corrections to the cumulants are given by

$$\kappa_r = \bar{\kappa}_r - \frac{B_r h^r}{r}\left(1 - \frac{1}{m^r}\right)$$

(Cf. Craig, 1936.)

3.14. Liapounoff's inequality for moments. Beginning with the inequality

$$(\Sigma ab)^2 \leqslant (\Sigma a^2)(\Sigma b^2)$$

show that for positive values $x_1 \ldots x_N$

$$\left(\Sigma x^{\frac{\alpha_1+\alpha_2}{2}}\right)^2 \leqslant \left(\Sigma x^{\frac{\alpha_1}{2}}\right)\left(\Sigma x^{\frac{\alpha_2}{2}}\right).$$

Hence that

$$(\Sigma x^{\Sigma(\alpha)/p})^p \leqslant (\Sigma x^{\alpha_1})(\Sigma x^{\alpha_2})(\ . \ . \ . \ )(\Sigma x^{\alpha_m})$$

is true when p is of form 2^m. Hence show that it is true for any integral p by noting that if 2^m is the smallest power of 2 greater than p we may take

$$_{p+1} = \alpha_{p+2} = \qquad \alpha_{2m} = \frac{\alpha_1 + \ . \ . \ . \ \alpha_p}{}$$

Hence, putting $p = a - c$, $\alpha_1 = \ . \ . \ . \ = \alpha_{a-b} = c$, $\alpha_{a-b-1} = \ . \ . \ . \ \alpha_{a-c} = a$, show that

$$\nu_b^{a-c} \leqslant \nu_c^{a-b}\nu_a^{b-c}.$$

(The inequality remains true for a continuous variate, as may be seen by considering limiting processes.)

3.15. Show that for the bivariate distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp - \frac{1}{2(1-\rho^2)}\left\{\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right\} dx_1\, dx_2 \qquad \infty \leqslant x_1, x_2 \leqslant \infty$$

all cumulants κ_{rs}, r, $s > 2$, vanish; and further, if

$$\lambda_{rs} = \frac{\mu_{rs}}{\sigma_1^r\sigma_2^s}$$

$$\lambda_{rs} = (r+s-1)\rho\lambda_{r-1,\, s-1} + (r-)(s-1)(1-\rho^2)\lambda_{r-2,\, s-2}$$

$$\lambda_{2r,\, 2s} = \frac{(2r)!(2s)!}{2^{r+s}} \sum_{j=0}^{t} \frac{(2\rho)^{2j}}{(r-j)!\,(s-j)!\,(2j)!}$$

$$\lambda_{2r+1,\, 2s+1} = \frac{(2r+1)!(2s+1)!}{2^{r+s}}\rho \sum_{j=0}^{t} \frac{(2\rho)^{2j}}{(r-j)!(s-j)!(2j+1)!}$$

$$\lambda_{2r,\, 2s+1} = \lambda_{2r+1,\, 2s} = 0,$$

where t is the smaller of r and s. In particular,

$$\lambda_{11} = \rho, \quad \lambda_{31} = 3\rho, \quad \lambda_{51} = 15\rho, \quad \lambda_{71} = 105\rho, \quad \lambda_{91} = 945\rho\,;$$
$$\lambda_{22} = (1+2\rho^2), \quad \lambda_{24} = 3(1+4\rho^2), \quad \lambda_{26} = 15(1+6\rho^2)$$
$$\lambda_{28} = 105(1+8\rho^2), \quad \lambda_{2,\,10} = 945(1+10\rho^2)\,;$$
$$\lambda_{33} = 3\rho(3+2\rho^2), \quad \lambda_{35} = 15\rho(3+4\rho^2),$$
$$\lambda_{44} = 3(3+24\rho^2+8\rho^4).$$

CHAPTER 4

CHARACTERISTIC FUNCTIONS

Moment- and Cumulant-Generating Functions

4.1. In the previous chapter we considered the characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF = \Sigma\mu_r'\frac{(it)^r}{r!} \tag{4.1}$$

as a moment-generating function. We have also

$$\psi(t) = \log \phi(t) = \Sigma\kappa_r\frac{(it)^r}{r!}, \tag{4.2}$$

$\psi(t)$ being known as the Cumulative Function. It generates the cumulants in the same way that the characteristic function generates the moments. If the moment of order r exists, $\phi(t)$ can be expanded in powers of t at least as far as the term in $(it)^r$, and so can $\psi(t)$.

Other functions can be constructed which generate the moments. For example, since for $|tx| < 1$

$$\frac{1}{1-tx} = \sum_{j=0}^{\infty}(tx)^j$$

we have the formal expansion

$$\int_{-\infty}^{\infty}\frac{dF}{(1-tx)} = \sum_{j=0}^{\infty} t^r\mu_r'. \tag{4.3}$$

Generally if a function $\zeta(t)$ can be expanded as a power series in t, $\Sigma a_j t^j$ we have, subject to existence (and convergence when the series is infinite),

$$\int_{-\infty}^{\infty}\zeta(tx)\, dF = \Sigma a_j t^j \mu_j'. \tag{4.4}$$

Since

$$(1+t)^x = \sum_{j=0}^{\infty}\frac{x^{[j]}}{j!}t^j$$

we have

$$(t) = \int_{-\infty}^{\infty}(1+t)^x\, dF = \sum_{j=0}^{\infty}\frac{t^j}{j!}\mu'_{[j]} \tag{4.5}$$

and thus $\omega(t)$ may be regarded as a factorial moment-generating function. We may also define a factorial cumulant-generating function

$$\log \omega(t) = \Sigma\frac{t^j}{j!}\kappa_{[j]}\,. \tag{4.6}$$

though this function has not come into general use.

4.2. The generation of moments is by no means the most important property of the characteristic function, and in this chapter we discuss some of the theorems which give it a fundamental place in statistical theory.

We recall, in the first instance, that $\phi(t)$ always exists, since

$$\left|\int_{-\infty}^{\infty} e^{itx}\, dF\right| \leqslant \int_{-\infty}^{\infty} | e^{itx} |\, dF$$
$$\int_{-\infty}^{\infty} dF = 1 \qquad (4.7)$$

so that the defining integral converges absolutely. Further, $\phi(t)$ is uniformly continuous in t and differentiable j times under the integral sign if the resulting expressions exist and are uniformly convergent, for which it is sufficient that ν_j exists. For then

$$| \phi^{(j)}(t) | = \left| \int_{-\infty}^{\infty} x^j e^{ixt}\, dF \right|$$
$$\leqslant \int_{-\infty}^{\infty} | x^j |\, dF = \nu_j. \qquad (4.8)$$

The Inversion Theorem

4.3. We now prove the fundamental theorem of the theory of characteristic functions, which will be called the Inversion Theorem, namely that the characteristic function uniquely determines the distribution function; more precisely, if $\phi(t)$ is given by (4.1) then

$$F(x) - F(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\frac{1 - e^{-ixt}}{it}dt \quad . \quad . \quad . \quad (4.9)$$

the integral being understood as a principal value, i.e. as

$$\lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \phi(t)\frac{1 - e^{-ixt}}{it}\, dt.$$

Further, if $F(x)$ is continuous everywhere and $dF = f(x)\, dx$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)e^{-ixt}\, dt \quad . \quad . \quad . \quad (4.10)$$

the integral, as before, being a principal value if there is not separate convergence at the limits. Equation (4.10) may be compared with the form

$$\phi(t) = \int_{-\infty}^{\infty} f(x)e^{ixt}\, dx, \quad . \quad (4.11)$$

the comparison exhibiting the kind of reciprocal relationship which exists between $f(x)$ and $\phi(t)$.

As a preliminary we require an integral due to Dirichlet. It is easy to show that

$$J = \int_0^{\infty} \frac{\sin x}{x}\, dx = \frac{\pi}{2}.$$

Putting

$$u_n = \int_{n\pi}^{(n+1)\pi} \left|\frac{\sin x}{x}\right| dx$$

we have

$$J = u_0 - u_1 + u_2 \ . \ . \ . + (-1)^r u_r + \ . \ . \ .$$

in which the terms decrease monotonically to zero in absolute value. Now let $H(x)$ be a positive decreasing function. Consider

$$I_p = \int_0^{\infty} H\left(\frac{x}{p}\right)\frac{\sin x}{x}\, dx = \Sigma\int_{n\pi}^{(n+1)\pi} (-1)^r H\left(\frac{x}{p}\right) \left|\frac{\sin x}{x}\right| dx. \quad . \quad . \quad (4.12)$$

Writing $H(+0)$ as the limit of $H(\varepsilon)$ as $\varepsilon \to 0$ (ε positive), we have, in virtue of the decreasing property of H, that any term in the series on the right in (4.12) is not greater than $u_n H(+0)$. Further, as the series alternates in sign, the difference between I_p and $\int_0^{n\pi} H\left(\frac{x}{p}\right) \frac{\sin x}{x}\, dx$ is less than $u_n H(+0)$, which tends to zero uniformly in n. Consequently I_p is uniformly convergent and we have

$$\lim_{p \to \infty} \int_0^\infty H\left(\frac{x}{p}\right)\frac{\sin x}{x}\, dx = \int_0^\infty \lim H\left(\frac{x}{p}\right)\frac{\sin x}{x}\, dx = \frac{\pi}{2} H(+0) \quad . \quad . \quad (4.13)$$

Similarly we have

$$\lim_{p \to \infty} \int_{-\infty}^0 H\left(\frac{x}{p}\right)\frac{\sin x}{x}\, dx = \frac{\pi}{2} H(-0) \qquad (4.14)$$

By a simple change of sign the results are seen to be true if $H(x)$ is a negative increasing function. It is therefore true of any function which can be expressed as the sum or difference of a positive decreasing function and a negative increasing function, and in particular of a frequency function or distribution function.

Adding (4.13) and (4.14) and writing $H(0)$ for $\frac{1}{2}\{H(+0) + H(-0)\}$ we have

$$\lim_{p \to \infty} \int_{-\infty}^\infty H\left(\frac{x}{p}\right)\frac{\sin x}{x}\, dx = \pi H(0) \quad . \quad . \quad . \quad . \quad (4.15)$$

and putting px for x in this expression,

$$\lim_{p \to \infty} \int_{-\infty}^\infty H(x)\, \frac{\sin px}{x}\, dx = \pi H(0), \quad . \quad . \quad . \quad . \quad (4.16)$$

the so-called Dirichlet integral. If $H(x)$ is continuous at $x = 0$ the value

$$\tfrac{1}{2}\{H(+0) + H(-0)\}$$

is of course the usual value $H(x = 0)$.

Now consider

$$J_c = \int_{-c}^{c} \phi(t)\, dt \int_0^X e^{-it\xi}\, d\xi. \quad . \quad . \qquad (4.17)$$

Putting
$$\phi(t) = \int_{-\infty}^\infty e^{itx}\, dF(x)$$

we have
$$J_c = \int_{-c}^{c} dt \left\{ \int_{-\infty}^\infty e^{itx}\, dF(x) \int_0^X e^{-it\xi}\, d\xi \right\}.$$

The product in curly brackets may be equated to the double integral

$$\int_0^X \int_{-\infty}^\infty e^{it(x-\xi)}\, dF(x)\, d\xi$$

which is evidently uniformly convergent. Making the transformation

$$y = x - \xi$$
$$z = \xi$$

we have

$$\int_0^X \int_{-\infty}^\infty e^{ity}\, dF_1(y + z)\, dz$$

and integrating with respect to z,

$$\int_{-\infty}^{\infty} e^{ity}\Big[F(y+z)\Big]_0^X dy$$
$$= \int^{\infty} e^{ity}\{F(y+X) - F(y)\}dy.$$

This also is uniformly convergent in t and hence, integrating under the integral sign with respect to t, we have

$$J_c = \int_{-\infty}^{\infty}\left[\frac{e^{ity}}{iy}\right]_{-c}^{c}\{F(y+X) - F(y)\}\,dy$$
$$= \int_{-\infty}^{\infty} 2\frac{\sin cy}{y}\{F(y+X) - F(y)\}\,dy \quad . \quad . \quad . \quad . \quad . \quad (4.18)$$

since $\cos cy$ is an even function.

Now (4.18) is a Dirichlet integral and we have, therefore,

$$\lim_{c\to\infty} J_c = 2\pi\{F(X) - F(0)\}. \quad . \quad . \quad . \quad . \quad . \quad (4.19)$$

Referring again to (4.17) we thus have, writing now x for X,

$$F(x) - F(0) = \frac{1}{2\pi}\lim_{c\to\infty}\int_{-c}^{c}\phi(t)\,dt\int_0^x e^{-it\xi}\,d\xi$$

and integrating with respect to ξ,

$$F(x) - F(0) = \frac{1}{2\pi}\lim_{c\to\infty}\int_{-c}^{c}\phi(t)\frac{1-e^{-ixt}}{it}\,dt,$$

which is the result stated in (4.9). It is to be remembered that in virtue of our convention in arriving at (4.16), $F(x)$ at a saltus is $\frac{1}{2}\{F(x+) + F(x-)\}$.

4.4. This expression may be thrown into an alternative form. From the definition of $\phi(t)$ it is seen that $\phi(t)$ and $\phi(-t)$ are conjugate quantities, and we thus have

$$R(t) = \tfrac{1}{2}\{\phi(t) + \phi(-t)\}$$
$$I(t) = \tfrac{1}{2}\{\phi(t) - \phi(-t)\},$$

R and I being the real and imaginary parts of $\phi(t)$. Thus

$$F(x) - F(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(t)\frac{1-e^{-ixt}}{it}\,dt$$

and by a change of sign in t,

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(-t)\frac{1-e^{ixt}}{it}\cdot dt$$
$$= \frac{1}{2}\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\frac{\phi(t)-\phi(-t)}{it}(1-\cos xt) + \frac{\phi(t)+\phi(-t)}{it}.i\sin xt\right\}dt$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{R(t)\sin xt + I(t)(1-\cos xt)}{t}\,dt \quad . \quad . \quad . \quad . \quad . \quad (4.20)$$

This integral is, of course, real.

4.5. If now $F(x)$ has a derivative $f(x)$ we have

$$f(x) = \frac{1}{2\pi}\frac{d}{dx}\lim_{c\to\infty}\int_{-c}^{c}\phi(t)\frac{1-e^{-ixt}}{it}\,dt$$

The integral being uniformly convergent in x, the differentiation can be carried out on the integrand, and we have

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)e^{-ixt}\,dt,$$

the integral being a principal value.

4.6. Consider now the expression

$$\frac{J_c}{2c} = \frac{1}{2c}\int_{-c}^{c} \phi(t)e^{-ixt}\,dt. \tag{4.21}$$

If the distribution function F has a derivative f, this is equal in the limit to

$$\lim_{c\to\infty} \frac{J_c}{2c} = \lim \frac{1}{2c} f(x) = 0$$

and consequently $\frac{J_c}{2c}$ tends to zero everywhere where $F(x)$ is continuous and differentiable, i.e. if the frequency-distribution is continuous.

If, however, the distribution is discontinuous, consider one point of discontinuity, say the frequency f_j at x_j. The contribution of this part of the frequency to $\phi(t)$ will be $f_j\,e^{itx_j}$, and thus the contribution to $\frac{J_c}{2c}$ will be

$$\frac{1}{2c}\int_{-c}^{c} f_j\,e^{itx_j}\,e^{-itx}\,dt$$
$$= \frac{1}{2c} f_j\left[\frac{e^{it(x_j-x)}}{i(x_j-x)}\right]_{-c}^{c}.$$

If $x \neq x_j$ this clearly tends to zero; but if $x = x_j$ it becomes

$$\frac{1}{2c}\int_{-c}^{c} f_j\,dt = f_j.$$

Thus the function $\frac{J_c}{2c}$ tends to f_j at $x = x_j$.

Hence, if $\frac{J_c}{2c}$ tends to zero at a point x, there is no discontinuity in the distribution function at that point; but if it tends to a positive number f_j, the distribution function is discontinuous at that point and the frequency is f_j. This gives us a criterion whether a given characteristic function represents a continuous distribution or not.

Example 4.1

We found in Example 3.10 that the characteristic function of the normal distribution

$$dF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\,dx \qquad -\infty \leqslant x \leqslant \infty$$

is

$$\phi(t) = e^{-\frac{1}{2}t^2\sigma^2}.$$

Suppose we are given such a function and require to find the distribution, if any, of which it is the characteristic function.

In the first place we note that the distribution, if any, is continuous. For

$$\frac{J_c}{2c} = \frac{1}{2c}\int_{-c}^{c} e^{-\frac{t^2\sigma^2}{2}} e^{-itx}\,dt.$$

The integral is less in modulus than $\int_{-c}^{c} e^{-\frac{t^2\sigma^2}{2}}\,dt$ which is less than $\int_{-}^{\infty} e^{-\frac{t^2\sigma^2}{2}}\,dt = \sqrt{2\pi}$

Thus $\frac{J_c}{2c} \to 0$, everywhere. We have then for the frequency function, if any,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\frac{-t^2\sigma^2}{2}} e^{itx}\,dt$$

$$= \frac{e^{-\frac{x^2}{2\sigma^2}}}{2\pi}\int_{-\infty}^{\infty} \exp - \tfrac{1}{2}\left(t\sigma - \frac{ix}{\sigma}\right)^2 dt.$$

This may be regarded as an integral in the complex plane along the line parallel to the real axis. Taking $t\sigma - \frac{ix}{\sigma}$ as the new variable in place of t, we find that the integral is in fact $\frac{1}{\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\zeta^2}\,d\zeta = \frac{\sqrt{2\pi}}{\sigma}$.

Thus

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

This is everywhere positive and $\int_{-\infty}^{\infty} dF$ converges. Hence it is in fact a frequency function with the given expression as a characteristic function.

Example 4.2

To find the frequency function, if any, for which

$$\phi(t) = e^{-|t|}.$$

We note that $\frac{J_c}{2c}$ tends to zero and that the distribution, if any, is continuous. We then have for $f(x)$, if it exists,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-|t|}e^{-itx}\,dt$$

$$= \frac{1}{2\pi}\int_{0}^{\infty} e^{-t-itx}\,dt + \int_{-\infty}^{0} e^{t-itx}\,dt$$

$$= \frac{1}{2\pi}\int_{0}^{\infty} e^{-t}(e^{itx} + e^{-itx})\,dt$$

$$= \frac{1}{\pi}\int_{0}^{\infty} e^{-t}\cos tx\,dt.$$

This may be evaluated by two partial integrations. We find

$$f(x) = \frac{1}{\pi}\Big[-e^{-t}\cos tx\Big]_0^{\infty} - \frac{x}{\pi}\int_0^{\infty} e^{-t}\sin tx\,dt$$

$$= \frac{1}{\pi} + \frac{x}{\pi}\Big[e^{-t}\sin tx\Big]_0^{\infty} - \frac{x^2}{\pi}\int_0^{\infty} e^{-t}\cos tx$$

$$= \frac{1}{\pi} - x^2 f(x).$$

Thus

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad -\infty \leqslant x \leqslant \infty.$$

As before, this function can represent a frequency function, and it is readily verified that $f(x)$ has, in fact, the required characteristic function.

Example 4.3

Does there exist a frequency function for which

$$\phi(t) = e^{it} ?$$

We have

$$\frac{J_c}{2c} = \frac{1}{2c}\int_{-c}^{c} e^{it-itx}\,dt = \frac{1}{2c}\int_{-c}^{c} e^{it(1-x)}\,dt.$$

If $1-x$ is not zero the integral is

$$\int_{-c}^{c} [\cos\{(1-x)t\} + i\sin\{(1-x)t\}]dt.$$

Since $\sin t$ is an odd function this is equal to

$$\int_{-c}^{c} \cos\{(1-x)t\}\,dt = \left[-\frac{\sin(1-x)t}{1-x}\right]_{-c}^{c}.$$

This does not converge, but it is bounded and hence $\frac{J_c}{2c} \to 0$.

If, however, $x = 1$, the integral is simply

$$\int_{-c}^{c} dt$$

and thus $\frac{J_c}{2c} = 1$.

Thus there is unit frequency at $x = 1$ and it is seen at once that this accounts for the whole of the frequency, so that there is no frequency elsewhere. The distribution thus consists of a unit at $x = 1$. This is otherwise evident from the consideration that $\log\phi(t) = it$, so that the second cumulant is zero and there is no dispersion.

Example 4.4

For what distribution, if any, are the cumulants given by $\kappa_r = (r-1)!$?

The series

$$\sum_{j=0}^{\infty}\kappa_j\frac{(it)^j}{j!} = \Sigma\frac{(it)^j}{j}$$

converges absolutely for $|t| < 1$ and is thus equal to $\psi(t)$ if such a function exists. We have

$$\psi(t) = \Sigma\frac{(it)^j}{j} = -\log(1 - it)$$

and thus

$$\phi(t) = \frac{1}{(1-it)}.$$

If the frequency function exists we have

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ixt}}{1-it}\,dt.$$

This integral may be evaluated by integrating the complex function $\frac{e^{-ixz}}{1-iz}$ round a contour consisting of the real axis and the infinite semicircle below that axis. The first part reduces

to the integral we are seeking. On the semicircle of radius R we have $z = R(\cos\theta + i\sin\theta)$ and the integrand becomes

$$\frac{\exp(-ixR\cos\theta + xR\sin\theta)}{1 - iR\cos\theta + iR\sin\theta}.$$

θ here lies between π and 2π and hence $\sin\theta$ is negative. Hence *if x is positive* the expression is less in modulus than

$$\frac{e^{-xR}}{R},$$

i.e. tends to zero as $R \to \infty$.

Now the function $\dfrac{e^{-ixz}}{1 - iz}$ has a pole within the domain of integration at $z = -i$ and the residue there is e^{-x}. Hence

$$\begin{aligned} f(x) &= \frac{1}{2\pi}.2\pi\, e^{-x} \\ &= e^{-} \qquad 0 \leqslant x \leqslant \infty. \end{aligned}$$

More generally, if $\kappa_r = p(r-1)!$, $p > 0$, it will be found that the residue of $\dfrac{e^{-ixz}}{(1 - iz)^p}$ is $\dfrac{x^{p-1}e^{-x}}{\Gamma(p)}$, so that the distribution is

$$f(x) = \frac{e^{-x}x^{p-1}}{\Gamma(p)}, \qquad 0 \leqslant x \leqslant \infty, \quad p > 0.$$

Example 4.5

For what distribution, if any, are all cumulants of odd order zero and those of even order a constant, say $2a$?

We have

$$\psi(t) = 2a\left\{\frac{(it)^2}{2!} + \frac{(it)^4}{4!} + \dots\right\}.$$

This series converges and

$$\psi(t) = 2a(\cos t - 1)$$

Hence
$$\phi(t) = e^{2a(\cos t - 1)}.$$

If we try to integrate $\int_{-\infty}^{\infty} e^{2a(\cos t-1)}\, e^{-itx}\,dt$ in the ordinary way we fail. Let us then look into the question of continuity of the distribution function.

We have

$$\begin{aligned} J_c &= e^{-2a}\int_{-c}^{c} e^{2a\cos t}\, e^{-ixt}\,dt \\ &= e^{-2a}\int_{-c}^{c}\sum_{j=0}^{\infty}\frac{(2a)^j}{j!}\cos^j t\, e^{-ixt}\,dt. \end{aligned}$$

The series is uniformly convergent and hence

$$\begin{aligned} J_c &= e^{-2a}\sum_{j=0}^{\infty}\int_{-c}^{c}\frac{(2a)^j}{j!}\cos^j t\, e^{-ixt}\,dt \\ &= e^{-2a}\sum_{j=0}^{\infty}\int_{-c}^{c}\frac{(2a)^j}{j!}\cos^j t\cos xt\,dt, \end{aligned}$$

since $\sin xt$ is an odd function.

Consider now the integral $\int_{-\infty}^{\infty} 2^j \cos^j t \cos xt\, dt$. By a well-known expansion

$$
\begin{aligned}
2^j \cos^j t \cos xt &= \tfrac{1}{2}(e^{it} + e^{-it})^j (e^{ixt} + e^{-ixt}) \\
&= \tfrac{1}{2}(e^{ixt} + e^{-ixt})\left\{e^{ijt} + \binom{j}{1} e^{(j-2)it} + \ldots + e^{-ijt}\right\}.
\end{aligned}
$$

The only part of this expression of present interest is the constant term, the others not contributing more than a finite amount to J_c. The coefficient of e^0 is zero unless x is integral in absolute value, and in that case is

$$
\frac{1}{2}\left\{\binom{j}{\frac{j-x}{2}} + \binom{j}{\frac{j+x}{2}}\right\} = \binom{j}{\frac{j-x}{2}}.
$$

Thus $\dfrac{J_c}{2c}$ tends to zero unless x is integral in absolute value, and in the latter case

$$
\frac{J_c}{2c} \to e^{-2a} \sum_{j=x}^{\infty} \frac{a^j}{j!} \binom{j}{\frac{j-x}{2}}.
$$

Thus, if x is even, say $2r$, the frequency at $x = \pm 2r$ is

$$
\begin{aligned}
&e^{-2a}\left\{\frac{a^{2r}}{(2r)!} + \frac{a^{2r+2}}{(2r+2)!}\binom{2r+2}{1} + \frac{a^{2r+4}}{(2r+4)!}\binom{2r+4}{2} + \ldots\right\} \\
&= e^{-2a}\left\{\frac{a^{2r}}{(2r)!} + \frac{a^{2r+2}}{(2r+1)!1!} + \frac{a^{2r+4}}{(2r+2)!2!} + \ldots\right\},
\end{aligned}
$$

and if x is odd the frequency at $x = 2r + 1$ is

$$
\frac{a^{2r+1}}{(2r+1)!} + \frac{a^{2r+3}}{(2r+2)!1!} + \frac{a^{2r+5}}{(2r+3)!2!} + \ldots\}.
$$

We may now verify that these frequencies account for the whole of the characteristic function and hence that all frequencies have been found.

Conditions for a Function to be a Characteristic Function

4.7. Any function which is not negative in its range of definition and which is integrable in the Stieltjes sense can be a frequency function; and any non-decreasing function which increases from 0 to 1 in its range of definition can be a distribution function. There are much more restrictive conditions to be obeyed before a given function can be a characteristic function.

In the first place, let us note that it is a necessary and sufficient condition for a function $\phi(t)$ to be a characteristic function that

$$
F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\frac{1 - e^{-ixt}}{it}\, dt
$$

shall (except for an additive constant $F(0)$) be a distribution function. This, however, is not a very helpful criterion in practice.

Looking to the definition of $\phi(t)$ as $\int_{-\infty}^{\infty} e^{itx}\,dF$, we see that necessary conditions for $\phi(t)$ to be a characteristic function are

(a) that $\phi(t)$ must be continuous in t,

(b) that $\phi(t)$ is defined in every finite t interval,

(c) that $\phi(0) = 1$,

(d) that $\phi(t)$ and $\phi(-t)$ shall be conjugate quantities,

(e) that $| \phi(t) | \leqslant \int_{-\infty}^{\infty} |e^{itx}|\,dF \leqslant 1 = \phi(0)$.

These criteria enable us to reject certain functions as possible characteristic functions, but there do not appear to be any readily applicable sufficient conditions which enable us to determine at sight whether a given function can be a characteristic function.

Limiting Properties of Distribution and Characteristic Functions

4.8. Suppose there is given a sequence of distribution functions $F_n(x)$ depending on a parameter n which can increase indefinitely. To each F_n there will correspond a characteristic function ϕ_n. The question to be discussed is this : if F_n tends to a limit F, will ϕ_n tend to a limit ϕ and is ϕ the characteristic function of F ? Conversely, if ϕ_n tends to a limit ϕ, does F_n tend to a limit F and is F a distribution function having ϕ for its characteristic function ? The answers to these questions, as will be seen below, are affirmative under certain general conditions.

It is to be noted what is meant by a distribution function tending to another. If both are continuous, $F_n(x)$ is said to tend to $F(x)$ if, given any ε, there is an n_0 such that $| F_n(x) - F(x) | < \varepsilon$ for all $n > n_0$. If there are discontinuities present, F_n will be said to tend to F if it does so in every point of continuity of F. Since by definition our functions are taken to be continuous on the left at saltuses, this evidently conforms to the definition for the continuous case and to the common-sense requirements of the situation.

4.9. We require two preliminary theorems for later work. The first is that if F_n tends to F it does so uniformly.

For the range can be divided into a finite number of parts, say at $\xi_1, \xi_2 \ldots \xi_h$, such that $F(\xi_{j+1}) - F(\xi_j) < \frac{\varepsilon}{2}$ for all j. Then as n increases there will come a time when $| F_n(\xi_j) - F(\xi_j) | < \frac{\varepsilon}{2}$ for all j. Thus there exists an n_0 such that for $n > n_0$

$$| F_n(\xi_j) - F(\xi_j) | < \frac{\varepsilon}{2}.$$

It is sufficient to show that this implies that for any x

$$| F_n(x) - F(x) | < \varepsilon, \qquad n > n_0.$$

In fact, if x lies between ξ_j and ξ_{j+1}

$$F(\xi_j) \leqslant F(x) \leqslant F(\xi_{j+1}) < F(\xi_j) + \frac{\varepsilon}{2}$$

and

$$F(\xi_j) - \frac{\varepsilon}{2} < F_n(\xi_j) \leqslant F_n(x) \leqslant F_n(\xi_{j+1}) < F(\xi_{j+1}) + \frac{\varepsilon}{2} < F(\xi_j) + \varepsilon$$

and thus $$-\varepsilon < F_n(x) - F(x) < \varepsilon,$$
which is the required result.

4.10. The second theorem we require (the Montel-Helly theorem) is that if the sequence $F_n(x)$ is monotonic and bounded for all x (which is so for distribution functions) then we can pick out a subsequence $F_{n'}(x)$ which converges to some monotonic increasing function F (not necessarily a distribution function itself, for it may not vary from 0 to 1).

Consider first of all a series of values $x_1, x_2, \ldots$ It is known that every bounded set of numbers contains a convergent sequence. Hence we can pick out from the sequence $F_n(x_1)$ a convergent sequence, say, $F_{n_1}(x_1)$. Then from the subsequence $F_{n_1}(x_2)$ we can pick out a subsequence $F_{n_2}(x_2)$ and $F_{n_2}(x)$ is thus convergent at both x_1 and x_2. Continuing in this way we may, by picking out the first function in $F_{n_1}(x)$, the second in $F_{n_2}(x)$, and so on, arrive at a sequence of functions $G_1(x)$, $G_2(x)$ . . . which converges at each of the values $x_1, x_2, \ldots$, etc. This is the so-called Weierstrassian diagonal process.

It follows that the sequence G_n is convergent at every *rational* point x. Since $G_n(a) \leqslant G_n(x) \leqslant G_n(b)$ for every x between a and b, we see that if $G_n(a)$ and $G_n(b)$ converge, the limiting values of $G_n(x)$ lie between those limits, say $G(a)$ and $G(b)$.

Then the function $u(x) =$ upper bound of $G_n(x)$ (x not necessarily rational) is well defined and non-decreasing and so has no more than an enumerable number of points of discontinuity. If u is continuous at x, we take y and z such that $y < x < z$ and $u(z) - u(y) < \varepsilon$. Then if a and b are rational points such that $y < a < x < b < z$ it follows that $u(y) \leqslant G(a) \leqslant G(b) \leqslant u(z)$. Moreover, as all the limiting values of $G_n(x)$ are between $G(a)$ and $G(b)$, they are between $u(y)$ and $u(z)$. Hence, as ε can be arbitrarily small, we see that $G(x)$ tends to $u(x)$ at every point of continuity of u. Finally, by the diagonal process, we can select a sequence which will also be convergent at the points of discontinuity of $u(x)$. The theorem is established.

The First Limit Theorem

4.11. We now prove the theorem: if a sequence of distribution functions F_n tends to a distribution function F, then the corresponding sequence of characteristic functions ϕ_n tends to ϕ uniformly in any finite t-interval, where ϕ is the characteristic function of F.

It is required to prove that, given ε, there is an n_0 independent of t such that

$$| \phi(t) - \phi_n(t) | = | \int_{-\infty}^{\infty} e^{itx}(dF - dF_n) | < \varepsilon, \qquad n > n_0.$$

Select two points of continuity of F, X and $-X$. We can make X as large as we please. We then split the integral

$$\int_{-c}^{\infty} e^{itx}(dF - dF_n) \qquad . \quad (4.22)$$

into two parts, that in the range $-X$ to $+X$ and that in the remaining portion of the range. Now

$$\int_{x<-X}^{x>X} e^{itx}\, dF \qquad \int_{x<-X}^{x>X} dF \leqslant 1 - F(X) - F(-X),$$

and by taking X large enough we can make this quantity less than $\frac{\varepsilon}{6}$.

Similarly

$$\int_{x<X}^{x>X} e^{itx}\, dF_n \leqslant 1 - F_n(X) - F_n(-X),$$

and since F_n tends to F (and that uniformly) this, for some large X, will be less than $\frac{\varepsilon}{3}$. Hence for some n_0 the portion of (4.22) outside the range $-X$ to $+X$ will be less in modulus than $\frac{\varepsilon}{6}+\frac{\varepsilon}{3}=\frac{\varepsilon}{2}$. Consider now the other part

$$\int_{-X}^{X} e^{itx}(dF-dF_n) \quad . \quad . \quad . \quad . \quad . \quad (4.23)$$

This expression is the limit of the sum

$$\Sigma e^{itx_j}[\{F(\xi_{j+1})-F(\xi_j)\}-\{F_n(\xi_{j+1})-F_n(\xi_j)\}], \quad . \quad . \quad . \quad (4.24)$$

ξ_{j}, ξ_{j+1} being the boundaries of the interval into which the range is subdivided and x_j a value in that interval. The difference between this sum and the limiting value can be made less than $\frac{\varepsilon}{4}$ if the intervals are small enough; for if they are less than η in width the difference of e^{itx_j} and $e^{it\xi_j}$ is less in modulus than $\eta\,|\,t\,|$, by the mean value theorem, and thus in any t-range $\pm T$ the difference of (4.23) and (4.24) is less in modulus than

$$\eta T\,|\,\Sigma[\{F(\xi_{j+1})-F(\xi_j)\}-\{F_n(\xi_{j+1})-F_n(\xi_j)\}]\,|<2\eta T,$$

which is less than $\frac{\varepsilon}{4}$ if $\eta<\frac{\varepsilon}{8T}$.

Now the sum (4.24) will itself be less than $\frac{\varepsilon}{4}$ for some $n>n_0$, for it is the sum of a finite number of terms each of which tends to zero.

Consequently (4.23) is less than $\frac{\varepsilon}{2}$ and hence

$$|\,\phi(t)-\phi_n(t)\,|<\varepsilon, \qquad n>n_0.$$

Converse of the First Limit Theorem

4.12. The converse result is even more important:

Let ϕ_n be a sequence of characteristic functions corresponding to the sequence of distribution functions F_n. Then if $\phi_n(t)$ tends to $\phi(t)$ uniformly in *some* finite t-interval, F_n tends to a distribution function F and ϕ is the characteristic function of F.

As a preliminary lemma, let us prove that if F is a distribution function with characteristic function ϕ, then for all real ξ and all $h>0$

$$\frac{1}{h}\int_{\xi}^{\xi+h}F(u)\,du-\frac{1}{h}\int_{\xi-h}^{\xi}F(u)\,du=\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2 e^{-\frac{2it\xi}{h}}\phi\left(\frac{2t}{h}\right)dh \quad . \quad . \quad (4.25)$$

In fact, put

$$G(x)=\frac{1}{h}\int_{x}^{x+h}F(u)\,du.$$

This is a continuous distribution function and its characteristic function is

$$\frac{1}{h}\int_{-\infty}^{\infty}e^{itx}\,dG$$

$$=\frac{1}{h}\int_{-\infty}^{\infty}e^{itx}\{F(x+h)-F(x)\}\,dx,$$

which by a partial integration becomes

$$\frac{1}{h}\left[\frac{\{F(x+h)-F(x)\}e^{itx}}{it}\right]_{-\infty}^{\infty}-\frac{1}{ith}\int_{-\infty}^{\infty}e^{itx}\{dF(x+h)-dF(x)\}\,dx$$

$$=\frac{-1}{ith}\int_{-\infty}^{\infty}\{e^{it(x-h)}\,dF(x)-e^{itx}\,dF(x)\}$$

$$=\phi(t)\frac{1-e^{-ith}}{ith}$$

Substituting for $G(x)$ in (4.9) we get

$$\frac{1}{h}\int_{\xi+h}^{\xi+2h}F(u)\,du-\frac{1}{h}\int_{\xi}^{\xi+h}F(u)\,du$$

$$=\frac{1}{2\pi h}\int_{-\infty}^{\infty}\left(\frac{1-e^{-iht}}{it}\right)^2e^{-it\xi}\,\phi(t)\,dt,$$

whence, writing ξ for $\xi+h$, we find

$$\frac{1}{h}\int_{\xi}^{\xi+h}F(u)\,du-\frac{1}{h}\int_{\xi-h}^{\xi}F(u)\,du=\frac{1}{2\pi h}\int_{-\infty}^{\infty}\left(\frac{1-e^{-ith}}{}\right)^2e^{-it(\xi-h)}\phi(t)\,dt$$

$$=\frac{1}{4\pi}\int_{-\infty}^{\infty}\left(\frac{e^{it}-e^{-it}}{it}\right)^2e^{-\frac{2it\xi}{h}}\phi\left(\frac{2t}{h}\right)dt$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2e^{-\frac{2it\xi}{h}}\phi\left(\frac{2t}{h}\right)dt,$$

the result announced in (4.25).

Reverting now to the theorem required to be proved, note that it is sufficient to establish that if $\phi_n\to\phi$ uniformly in *some* interval $|t|<a$, then F_n tends to some distribution function F in every point of continuity of F. When this is established it follows from the First Limit Theorem that ϕ is the characteristic function of F and that ϕ_n converges to ϕ uniformly in *every* finite t-interval.

As shown in **4.10**, given a sequence F_n we may always choose from it a subsequence $F_{n'}$ such that $F_{n'}$ converges to a non-decreasing function F in every continuity point of F.

Let us then choose such a sequence. We have of necessity $0\leqslant F\leqslant 1$, and F may be supposed everywhere continuous on the left. It is then a distribution function if $F(+\infty)-F(-\infty)=1$, and this we proceed to prove.* From (4.25) with $\xi=0$ we have

$$\frac{1}{h}\int_0^h F_{n'}(u)\,du-\frac{1}{h}\int_{-h}^0F_{n'}(u)\,du=\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2\phi_{n'}\left(\frac{2t}{h}\right)dt.$$

By hypothesis ϕ_n tends uniformly to ϕ for $|t|<a$ and hence $\phi_{n'}$ does so, and it is easily seen that the integral on the right is uniformly convergent. Thus, given ε, we can find h_0 such that for $h>h_0$

$$\frac{1}{h}\int_0^h-\frac{1}{h}\int_{-h}^0F(u)\,du=\frac{1}{\pi}\int_{-\frac{1}{2}ah}^{\frac{1}{2}ah}\left(\frac{\sin t}{t}\right)^2\phi\left(\frac{2t}{h}\right)dt+\eta,$$

* It is not obvious that if the functions F_n all vary from 0 to 1, then their limit must do so. In fact, if $F_n(x)=0,\ x<-n,\quad F_n(x)=\frac{1}{2},\ -n\leqslant x\leqslant n,\quad F_n(x)=1,\ x>n$ then $\lim_{n\to\infty}F_n(x)=\frac{1}{2}$ for $-\infty\leqslant x\leqslant\infty$.

where $|\eta| < \varepsilon$. Now let h tend to infinity. As F is a non-decreasing function the left-hand side tends to $F(+\infty) - F(-\infty)$. The right-hand side tends, in virtue of the uniformity of $\phi_{n'}$ and the consequent continuity of ϕ near $t = 0$, to

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2 dt,$$

which is equal to unity.

Hence F, the limit of the subsequence $F_{n'}$, is a distribution function whose characteristic function is ϕ.

But any subsequence of ϕ_n tends to ϕ, in virtue of the uniformity of the convergence, and hence any convergent subsequence of F_n tends to F. Consequently F_n tends to F in every point of continuity of F and the theorem follows.

Example 4.6

The binomial distribution $(q + p)^n$ considered in Example 3.2 has the characteristic function

$$(q + pe^{it})^n.$$

Now the frequency at $x = j$ is $\binom{n}{j}q^{n-j}p^j$. This is greater than the ordinate at $x = j + 1$ if

$$\binom{n}{j}q^{n-j}p^j > \binom{n}{j+1}q^{n-j-1}p^{j+1}$$

or

$$j > pn - q.$$

For large n the maximum frequency will then be in the neighbourhood of $j = pn$, and is then

$$n \quad q^{qn}p^{pn}.$$

In virtue of Stirling's approximation to the factorial this approximates to

$$\frac{n^n e^{-n}\sqrt{2\pi n}\, q^{qn}p^{pn}}{(pn)^{pn}e^{-pn}\sqrt{(2\pi pn)}(qn)^{qn}e^{-qn}\sqrt{(2\pi qn)}} \sim \frac{1}{\sqrt{(2\pi pqn)}},$$

and therefore tends to zero.

Thus every frequency in the binomial tends to zero and the distribution does not tend to any limiting distribution.

Suppose, however, that we express the distribution in standard measure. Putting $\xi = \frac{x-\mu_1'}{}$ we have

$$\phi_x(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x) = \int_{-\infty}^{\infty} e^{it(\sigma\xi+\mu_1')}\, dF(\xi)$$

$$= e^{it\mu_1'}\phi_\xi(\sigma t).$$

Hence

$$\phi_\xi(t) = e^{-\frac{it\mu_1'}{\sigma}}\phi_x\left(\frac{t}{\sigma}\right).$$

The effect on $\phi(t)$ of transferring to standard measure is then to replace t by $\frac{t}{\sigma}$ and to multiply by $e^{-\frac{it\mu_1'}{\sigma}}$.

For the binomial $\mu_1' = np$, $\mu_2 = npq$, and thus the characteristic function of the binomial expressed in standard measure is

$$\exp\left(-\frac{itnp}{(npq)^{\frac{1}{2}}}\right)\left(q + p\exp\frac{it}{(npq)^{\frac{1}{2}}}\right)^n$$

Thus

$$\begin{aligned}\log\phi &= \frac{-itnp}{(npq)^{\frac{1}{2}}} + n\log\left\{1 + p\left(\exp\frac{it}{(npq)^{\frac{1}{2}}} - 1\right)\right\}\\ &= \frac{-itnp}{(npq)^{\frac{1}{2}}} + n\log\left\{1 + p\frac{it}{(npq)^{\frac{1}{2}}} \quad \frac{pt^2}{2npq} + \frac{p\theta t^3}{6(npq)^{\frac{3}{2}}}\right\} \qquad 0 \leqslant |\theta| \leqslant 1\\ &= n\left\{\frac{-pt^2}{2npq} - \frac{1}{2}\frac{(pit)^2}{npq}\right\} + 0(t^3n^{-\frac{1}{2}})\\ &= -\tfrac{1}{2}t^2 + 0(t^3n^{-\frac{1}{2}}).\end{aligned}$$

Thus for any finite t $\log\phi$ tends uniformly to $-\frac{1}{2}t^2$ and hence

$$\phi(t) \rightarrow e^{-\frac{1}{2}t^2}.$$

Thus the distribution $(q+p)^n$ expressed in standard measure tends to the distribution whose characteristic function is $e^{-\frac{1}{2}t^2}$, i.e. to the form

$$dF = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Multivariate Characteristic Functions

4.13. The characteristic function of a bivariate distribution $F(x_1, x_2)$ is defined as

$$\phi(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it_1x_1+it_2x_2}\,dF(x_1, x_2) \qquad . \quad . \quad . \quad (4.26)$$

and generally, that of a multivariate distribution $F(x_1, x_2, \ldots x_n)$ as

$$\phi(t_1, t_2, \ldots t_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} e^{it_1x_1+it_2x_2+\ldots it_nx_n}\,dF(x_1, x_2, \ldots x_n) \qquad (4.27)$$

If $x_1, x_2, \ldots x_n$ are independent we have

$$\begin{aligned}\phi(t_1, t_2, \ldots t_n) &= \int_{-\infty}^{\infty} e^{it_1x_1}\,dF_1(x_1)\int_{-\infty}^{\infty} e^{it_2x_2}\,dF_2(x_2) \ldots \int_{-\infty}^{\infty} e^{it_nx_n}\,d_rF_n(x_n)\\ &= \phi(t_1)\phi(t_2)\ldots\phi(t_n) \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.28)\end{aligned}$$

Similarly

$$\psi(t_1, t_2 \ldots t_n) = \sum_{j=1}^{n}\log\phi(t_j) \qquad . \quad . \quad . \quad . \quad (4.29)$$

Thus the characteristic function of the joint distribution of a number of independent variables is the product of their characteristic functions; and the cumulative function is the sum of their cumulative functions. This is a fundamentally important result in the theory of sampling.

4.14. In generalisation of (4.9) we have

$$F(x_1, x_2, \ldots x_n) = \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\frac{1-e^{ix_1t_1}}{it_1} \quad \frac{1-e^{ix_nt_n}}{it_n}$$
$$\phi(t_1, t_2, \ldots t_n)\,dt_1\ldots dt_n. \qquad . \qquad . \quad (4.30)$$

The multiple integrals are to be interpreted as principal values

$$\lim_{c \to \infty} \int_{-c}^{c} \dots \int_{-c}^{c}.$$

The proof is similar to that for the univariate case. We have

$$J = \int_0^\infty \dots \int_0^\infty \frac{\sin x_1}{x_1} \dots \frac{\sin x_n}{x_n} dx_1 \dots dx_n = \left(\frac{\pi}{2}\right)^n$$

$$\lim_{p \to \infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} H\left(\frac{x_1}{p}, \dots \frac{x_n}{p}\right) \frac{\sin x_1}{x_1} \dots \frac{\sin x_n}{x_n} dx_1 \dots dx_n = \pi^n H(+0 \dots +0)$$

$$\lim_{p \to \infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} H(x_1, \dots x_n) \frac{\sin px_1}{x_1} \dots \frac{\sin px_n}{x_n} dx_1 \dots dx_n = \pi^n H(0, \dots 0).$$

Considering now

$$J_c = \int_{-c}^{c} \dots \int_{-c}^{c} \phi(t_1, \dots t_n) dt_1 \dots dt_n \int_0^{X_1} \dots \int_0^{X_n} \exp(-it_1\xi_1 \dots - it_n\xi_n) d\xi_1 \dots d\xi_n \quad (4.31)$$

we find that

$$J_c = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{2 \sin cx_1}{x_1} \dots \frac{2 \sin cx_n}{x_n} \{F(X_1 + x_1, \dots X_n + x_n) - F(x_1, \dots x_n)\} dx_1 \dots dx_n.$$

$$\lim_{c \to \infty} J_c = (2\pi)^n \{F(x_1, \dots x_n) - F(0, \dots 0)\}$$

and by considering the integration of (4.31) with respect to the ξ's the result (4.30) follows.

4.15. If we have a distribution $F(x)$ and some function of the variate such as $\xi(x)$ we may consider the characteristic function of ξ

$$\phi_\xi(t) = \int_{-\infty}^{\infty} e^{it\xi} \, dF(x) \quad . \quad (4.32)$$

The distribution of ξ will then be given by (4.9) or (4.10), e.g. the distribution function of ξ, say $G(\xi)$, is

$$G(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-it\xi) \phi_\xi(t) \, dt. \quad . \quad (4.33)$$

The Problem of Moments

4.16. We can now consider in more detail a problem which suggested itself in Chapter 3. Do the moments determine the distribution uniquely, and if not, under what conditions do they do so? To give some point to this question let us note that in some circumstances it is possible for two different distributions to have the same set of moments.

Consider in fact the integral

$$\int_0^\infty t^{p-1} e^{-qt} dt = \frac{\Gamma(p)}{q^p}, \qquad p > 0, \ q > 0.$$

Put $\quad p = \dfrac{(n+1)}{\lambda} \qquad n$ a non-negative integer

$$0 < \lambda < \tfrac{1}{2}$$

$$q = \alpha + i\beta$$

$$\frac{\beta}{\alpha} = \tan \lambda\pi$$

$$x^\lambda = t.$$

We find on substitution that

$$\int_0^\infty x^n e^{-\alpha x^\lambda}\{\cos(\beta x^\lambda \tan\lambda\pi) + i\sin(\beta x^\lambda \tan\lambda\pi)\}\,\lambda\,dx\,. \qquad (4.34)$$

$$= \frac{\Gamma\left(\frac{n+1}{\lambda}\right)}{\alpha^{\frac{n+1}{\lambda}}(1+i\tan\lambda\pi)^{\frac{n+1}{\lambda}}}$$

and since $$(1+i\tan\lambda\pi)^{\frac{n+1}{\lambda}} = \frac{\cos(n+1)\pi + i\sin(n+1)\pi}{(\cos\lambda\pi)^{\frac{n+1}{\lambda}}}$$

$$= \text{a real quantity},$$

the imaginary part of (4.34) is zero. Thus the distributions

$$f(x) = ke^{-\alpha x^\lambda}\{1 + \varepsilon\sin(\beta x^\lambda \tan\lambda\pi)\} \quad . \quad . \quad . \quad . \quad (4.35)$$

$$0 \leqslant x \leqslant \infty, \quad \alpha > 0, \quad 0 < \lambda < \tfrac{1}{2}, \quad |\varepsilon| < 1$$

have moments independent of ε, and (4.35) defines a whole family of distributions having the same moments.

Similarly, if we substitute

$$p = \frac{(2n+1)}{\lambda}, \quad q = \alpha + i\beta, \quad \frac{\beta}{\alpha} = \tan\frac{\rho\pi}{2}, \quad x^\rho = t, \quad \rho = \frac{2s}{s+1} (s \text{ a positive integer})$$

we find that the family

$$f(x) = ke^{-\alpha x^\rho}\left\{1 + \varepsilon\cos\left(\alpha x^\rho \tan\frac{\rho\pi}{2}\right)\right\} \quad . \quad . \quad . \quad . \quad (4.36)$$

$$-\infty \leqslant x \leqslant \infty, \quad \alpha > 0, \quad 0 < \rho = \frac{2s}{s+1} < 1, \quad |\varepsilon| < 1$$

all have the same moments, the range in this case being infinite in both directions.

4.17. In full generality the problem of moments may be formulated as follows: Given a sequence of constants $c_0, c_1, \ldots c_j \ldots$,

(i) Does there exist a distribution function F such that

$$\int_a^b x^r\,dF = c_r\,? \quad . \quad (4.37)$$

(ii) If so, is the distribution function unique?

(iii) What are the functions, if any?

We have not the space here to enter on a full discussion of these questions, which have stimulated some beautiful mathematics, particularly by Stieltjes (1918). Our treatment will be confined to the results of statistical interest, but we may indicate the principal results of Stieltjes.

If we express the series

$$\sum_{j=0}^{\infty}(-1)^j \frac{c_j}{z^j} \quad . \quad (4.38)$$

as a continued fraction of the form

$$\frac{1}{a_1 z +}\,\frac{1}{a_2 +}\,\frac{1}{a_3 z +}\,\frac{1}{a_4 +}\cdots\frac{1}{a_{2n-1}z +}\,\frac{1}{a_{2n} +} \quad . \quad . \quad . \quad (4.39)$$

then, if the limits in (4.37) are 0 to ∞, it is a necessary and sufficient condition for the

existence of at least one F that all the a's be positive; and F is unique or not according as $\sum_{j=1}^{\infty} a_j$ diverges or converges.

The case when the limits in (4.37) are $\pm\infty$ has been treated by Hamburger (1920), who showed that an F exists if the expression of (4.38) as a continued fraction of the form

$$\frac{b_1}{a_0 + z +}\ \frac{b_2}{a_1 + z +}\ a_2 + z + \quad . \quad (4.40)$$

gives positive values of the b's. In order that F may be unique it is necessary and sufficient that the continued fraction be completely convergent in a sense defined by Hamburger.

We shall see presently that for finite limits in the integral of (4.37) the function F is always unique.

4.18. Unfortunately the Stieltjes-Hamburger criteria are not of much practical use because, as a rule, it is too difficult to express the a's and b's of (4.39) and (4.40) explicitly enough in terms of the given c's to enable questions of sign or convergence to be decided. We may, however, derive some criteria of statistical importance by considering the more restricted problem: given the moments of a distribution, can any other distribution also have the moments? In other words, we are given the existence of one F and require to know whether F is unique.

Note in the first instance that this problem need only be considered when absolute moments of all orders exist. It is evident that more than one distribution can exist having a limited number of moments finite and the remainder infinite. Furthermore, if any moment of even order exists, those of lower order must exist. In particular, if μ_{2r} exists $\int_0^\infty x^{2r}\,dF$ and $\int_{-\infty}^0 x^{2r}\,dF$ exist separately, and hence so do $\int_0^\infty |x^{2r-1}|\,dF$ and $\int_{-}^0 |x^{2r-1}|\,dF$, and so also does $\int_{-\infty}^\infty |x^{2r-1}|\,dF$, the absolute moment of order $2r-1$. Thus we consider only the case when all absolute moments exist.

4.19. We will prove in the first place the theorem that a set of moments determines a distribution uniquely if the series $\sum_{j=0}^{\infty}\frac{\nu_j t^j}{j!}$ converges for some real non-zero t.

The characteristic function is continuous in t and its derivatives exist if the moments exist. We have then in the neighbourhood of $t = 0$

$$\phi(t) = \sum_{j=0}^{r}\frac{(it)^j}{j!}\mu'_j + R_r \quad . \quad (4.41)$$

where R_r is less in absolute value than $\frac{3\nu_r t^r}{r!}$ (3.14).

Thus if $\frac{\Sigma\nu_j t^j}{j!}$ converges, $\frac{\nu_j t^j}{j!}$ tends to zero and hence $\phi(t)$ is equal to the sum of the infinite series $\sum_{j=0}^{\infty}\frac{(it)^j\mu'_j}{j!}$ if it exists. Moreover, this series is majorated by $\frac{\Sigma\nu_j t^j}{j!}$ and hence

is absolutely convergent if the latter is convergent. Hence we have

$$\phi(t) = \sum_{j=0}^{\infty} \frac{(it)^j \mu'_j}{j!} \quad . \quad . \quad . \quad . \quad . \quad (4.42)$$

and thus $\phi(t)$ is uniquely determined in the neighbourhood of $t = 0$. In the neighbourhood of $t = t_0$ we have

$$\phi(t) = \sum_{j=0}^{r} \left\{ \frac{i^j (t - t_0)^j}{j!} \int_{-\infty}^{\infty} x^j e^{it_0 x}\, dF \right\} + R'_r$$

and the modulus of the coefficient of $\frac{(t - t_0)^j}{j!}$ is not greater than ν_j. Consequently $\phi(t)$ can be expanded everywhere as a convergent Taylor series and is equal to the sum of that series. Hence $\phi(t)$ may be extended from the neighbourhood $t = t_0$ by analytic continuation through any finite t-interval. Hence $\phi(t)$ is everywhere uniquely defined.

But $\phi(t)$ determines the distribution function and hence the latter is uniquely determined.

4.20. As a corollary of this theorem we have the result that a set of moments uniquely determines a distribution if

$$\overline{\lim}\ \frac{\nu_n^{\frac{1}{n}}}{n} \text{ is finite.} \qquad (4.43)$$

For the series $\Sigma \frac{\nu_n t^n}{n!}$ is convergent if

$$\overline{\lim} \left(\frac{\nu_n t^n}{n!} \right)^{\frac{1}{n}} < 1,$$

that is to say, in virtue of the Stirling approximation to the factorial, if

$$\lim \left(\frac{\nu_n e^n t^n}{n^n} \right)^{\frac{1}{n}} = \overline{\lim}\ \frac{\nu_n^{\frac{1}{n}}}{n} et < 1.$$

If k is the upper limit of $\frac{\nu_n^{\frac{1}{n}}}{n}$ this inequality will be satisfied for $t < \frac{1}{ek}$.

Is is also a sufficient condition for uniqueness that $\overline{\lim}\ \frac{\mu'^{\frac{1}{2n}}_{2n}}{h}$ should be finite, a form of the criterion which enables us to disregard the *absolute* moments. In fact

$$\nu_{2n-1}^{\frac{1}{2n-1}} \leqslant \nu_{2n}^{\frac{1}{2n}} = \mu'^{\frac{1}{2n}}_{2n} \leqslant \nu_{2n+1}^{\frac{1}{2n+1}}$$

so that

$$\frac{1}{2n-1} \nu_{2n-1}^{\frac{1}{2n-1}} \leqslant \frac{2n}{2n-1} \cdot \frac{1}{2n} \mu'^{\frac{1}{2n}}_{2n} \leqslant \frac{2n+1}{2n-1} \frac{1}{2n+1} \nu_{2n+1}^{\frac{1}{2n+1}}$$

Taking upper limits throughout we have

$$\overline{\lim}\ \frac{1}{2n-1} \nu_{2n-1}^{\frac{1}{2n-1}} \leqslant \overline{\lim}\ \frac{1}{2n} \mu'^{\frac{1}{2n}}_{2n} \leqslant \overline{\lim}\ \frac{1}{2n+1} \nu_{2n+1}^{\frac{1}{2n+1}}$$

and thus $\overline{\lim}\ \frac{1}{n} \nu_n^{\frac{1}{n}}$ and $\overline{\lim}\ \frac{1}{2n} \mu'^{\frac{1}{2n}}_{2n}$ are finite or infinite together.

4.21. As a further corollary we have the result that a set of moments uniquely determines a distribution if the range is finite. For suppose the range is a to b. Taking an origin at $x = a$ and letting $b - a = c$, we have

$$\mu_n' = \int_a^b x^n \, dF \leqslant c^n.$$

Thus
$$\mu_n'^{\frac{1}{n}} \leqslant c = \nu_n^{\frac{1}{n}}$$

and hence $\overline{\lim} \dfrac{\nu_n^{\frac{1}{n}}}{n} = 0.$

4.22. Two further criteria may be mentioned. The first is due to Carleman (1925). A set of moments determines a distribution uniquely if (in the case of limits $-\infty$ to $+\infty$)

$$\sum_{j=0}^{\infty} \frac{1}{(\mu_{2j})^{\frac{1}{2j}}} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.44)$$

diverges. For the limits 0 to ∞ the corresponding series is

$$\sum_{j=0}^{\infty} \frac{1}{(\mu_j)^{\frac{1}{2j}}} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.45)$$

Secondly, if there exists a frequency function, the moments determine it uniquely if, for limits $-\infty$ to $+\infty$

$$f(x) < M \,|\, x \,|^{\beta-1} e^{-\alpha|x|^{\lambda}} \text{ for } |\, x \,| > x_0 \;\; (M, \beta, \alpha, > 0 \quad \lambda \leqslant 1) \quad . \quad . \quad (4.46)$$

and for limits 0 to ∞

$$f(x) < M \,|\, x \,|^{\beta-1} e^{-\alpha|x|^{\lambda}} \text{ for } |\, x \,| > x_0 \;\; (M, \beta, \alpha, > 0 \quad \lambda \leqslant \tfrac{1}{2}) \quad . \quad . \quad (4.47)$$

This result is due ultimately to Stieltjes. It follows without difficulty from the Carleman criterion.

It is interesting to note that if for some x_0

$$f(x) > e^{-\alpha|x|^{\lambda}} \qquad (\alpha > 0) \qquad . \quad . \quad . \quad . \quad . \quad (4.48)$$

then the problem of moments is necessarily indeterminate (as usual, $\lambda < \frac{1}{2}$ for the range 0 to ∞ and $\lambda < 1$ for the range $-\infty$ to $+\infty$). This follows from the examples in equations (4.35) and (4.36), for we can add to (4.48), *without rendering any frequency negative*, a function all of whose moments are zero.

Example 4.7

The moments of the distribution

$$dF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \, dx, \qquad -\infty \leqslant x \leqslant \infty$$

are given by
$$\mu_{2r+1} = 0$$
$$\mu_{2r} = \frac{(2r)!}{2^r r!} \sigma^{2r}$$

Thus
$$\begin{aligned}\frac{1}{n}\,\mu_{2n}^{\frac{1}{2n}} &\sim \frac{\sigma}{n\sqrt{2}}\left\{\frac{(2n)!}{n!}\right\}^{\frac{1}{2n}}\\ &\sim \frac{\sigma}{n\sqrt{2}}\left\{\frac{e^{-2n}(2n)^{2n}\sqrt{(4\pi n)}}{e^{-n}n^{n}\sqrt{(2\pi n)}}\right\}^{\frac{1}{2n}}\\ &\sim \frac{\sigma}{n\sqrt{(2e)}}\,\frac{2n}{n^{\frac{1}{2}}}\\ &\sim \frac{\sigma}{\sqrt{(2e)}n^{\frac{1}{2}}}\end{aligned}$$
and thus the upper limit is zero and the distribution is unique.

4.23. If the moment of order r exists it must be given by
$$i^r\mu_r' = \left[\frac{d^r}{dt^r}\,\phi(t)\right]_{t=0}.$$
Thus if $\phi(t)$ can be expanded in an infinite Taylor series, that series must be $\Sigma\,\frac{(it)^r}{r!}\mu_r'$. Further, if this series does not converge, $\phi(t)$ cannot be expanded as an infinite Taylor series. But it can always be expanded in the finite form with remainder
$$\phi(t) = \sum_{j=0}^{r}\frac{(it)^j}{j!}\mu_j' + R.$$
Thus, when the series does not converge, $\phi(t)$ can be expanded in powers of t only asymptotically.

This illustrates the source of the ambiguity in the definition of $\phi(t)$ when the infinite series $\Sigma\frac{(it)^j}{j!}\mu_j'$ does not converge, for it is known that there exist an infinite number of functions which have a given set of coefficients in an asymptotic expansion. For instance, if $\alpha(t)$ has an asymptotic expansion in t the functions $\alpha(t) + kt^{-\log t}$ all have the same expansion. It is therefore hardly surprising that when $\Sigma\frac{\nu_r t^r}{r!}$ or $\Sigma\,\frac{(it)^r}{r!}\,\mu_r'$ fail to converge, there may be more than one frequency or distribution function with the same set of moments.

But it does not follow from what has been said that there *must* be more than one frequency-distribution. There must be more than one function, but those functions may not qualify as frequency-distributions, e.g. they may be negative in part of the range. In the example just given, $t^{-\log t}$ cannot be a characteristic function, for it does not obey the well-known condition that $\phi(t)$ and $\phi(-t)$ should be conjugate. So far as I am aware, it is not known whether the condition that $\Sigma\frac{(it)^j}{j!}\mu_j'$ should converge is necessary as well as sufficient for uniqueness.

The Second Limit Theorem

4.24. We are now in a position to prove a further theorem on the limits of distribution functions. If a sequence of functions $F_n(x)$ has all moments existing and for all j $\mu_j'(n) \rightarrow \mu_j'$, then the μ_j''s are the moments of a distribution function F which is the limit of the sequence F_n, provided that F is completely determined by its moments.

We will first prove the rather more general theorem : If there is given a sequence of distribution functions F_n such that all moments of F_n exist, and for any j the sequence $\mu_j'(n)$ lies between fixed limits independent of n, then a subsequence $F_{n'}$ can be selected from F_n such that

(1) $\lim\limits_{n\to\infty}\int_{-\infty}^{\infty} x^j\,dF_{n'}$ exists, $= \mu_j'$ say.

(2) The subsequence $F_{n'}$ converges to some distribution function F.

(3) $\int_{-\infty}^{\infty} x^j\,dF$ exists and is equal to μ_j'.

The existence of μ_j' may be proved by the diagonal method exactly as for the Montel-Helly theorem of **4.10.** By hypothesis, $\mu_j'(n)$ is uniformly bounded and the rest of the proof follows that of **4.10.**

The existence of F follows also from the Montel-Helly theorem. We apply the theorem to the subsequence derived by satisfying condition (1) and hence arrive at a subsequence obeying both (1) and (2). It must however be shown that F is a distribution function, i.e. varies effectively from 0 to 1. This follows because

$$\left.\begin{aligned}\int_b^{\infty} x^r\,dF_n &\leqslant \frac{\mu_{2r+2}'(n)}{b^{r+2}}, \qquad b>1\\ \int_{-\infty}^{a} x^r\,dF_n &\leqslant \frac{\mu_{2r+2}'(n)}{|a|^{r+2}} \qquad a<-1\end{aligned}\right\} \tag{4.49}$$

and hence, for the subsequence, with $r = 0$ and letting n' tend to infinity,

$$0 \leqslant 1 - F(b) \leqslant \frac{1+\mu_2'}{b^2}$$

$$0 \leqslant F(a) \leqslant \frac{1+\mu_2'}{a^2}$$

so that, as a, b tend to infinity the equations $F(\infty) = 1$, $F(-\infty) = 0$ are seen to hold.

We also require for later parts of the proof two results : the first that the convergence of

$$\lim_{a\to-\infty,\, b\to\infty}\int_a^b x^j\,dF_n \text{ to } \int_{-\infty}^{\infty} x^j\,dF_n \quad . \quad . \tag{4.50}$$

is uniform with respect to n. This follows from the hypothesis that $\mu_{(2r+2)}'(n)$ is bounded and from the equations (4.49). The second is that

$$\lim_{x\to\infty} x^s\{1 - F_n(x)\} = 0, \qquad \lim_{x\to-\infty} x^s \mid F_n(x) = 0 \quad . \quad . \tag{4.51}$$

for $s > 0$ and all integral $n > 0$. The first limit follows from

$$1 - F_n(b) = \int_b^{\infty} dF_n \leqslant \int_b^{\infty}\left(\frac{x}{b}\right)^{2j} dF_n$$

and hence from

$$b^s\{1 - F_n(b)\} \leqslant \frac{\mu_{2j}'}{b^{2j-s}}, \qquad b > 0, \quad 0 < s < 2j.$$

We now have to complete the proof by showing that $\int_{-\infty}^{\infty} x^j\,dF$ exists and is equal

to μ_j'. For this we use the theorem (an extension by Fréchet and Shohat (1931) of one due to Helly) that if a sequence $v_n(x)$, defined in the interval $-\infty$ to $+\infty$, is such that

(1) $v_n(x)$ is of bounded variation in any finite interval,

(2) all $v_n(x)$ and their total variations are bounded in any finite interval,

(3) $\lim\limits_{n\to\infty} v_n(x) = v(x)$ exists for all x, except perhaps at a denumerable number of points,

(4) $\int_a^b f(x)\,dv_n(x)$ converges uniformly with respect to n to $\int_{-\infty}^{\infty} f(x)\,dv(x)$ if $f(x)$ is everywhere continuous,

then $\int_{-\infty}^{\infty} f(x)\,dv(x)$ exists and $= \lim\limits_{n\to\infty}\int_{-\infty}^{\infty} f(x)\,dv_n(x)$.

This result may be applied to our sequence $F_n(x)$, which obeys conditions (1), (2) and (3). It also obeys (4) when $f(x) = x^j$ in virtue of (4.50) and (4.51). Further $F(x)$ is of bounded variation and hence $\int_{-\infty}^{\infty} x^j\,dF(x)$ exists and equals, say, μ_j'.

Finally

$$\mu_j' - \mu_j'(n)\,| = \Big|\;\; x^j(dF - dF_n)$$

$$\int_{-\infty}^{a} x^j\,dF \;+\; \int_{-\infty}^{a} x^j\,dF_n \qquad\qquad x^j\,dF$$

$$+ \qquad x^j\,dF_n \;+\; \int_a^b x^j\,(dF - dF_n) \qquad\qquad a < 0,\ b > 0. \qquad . \quad (4.52)$$

By taking $-a$ and b sufficiently large we can make the first four terms on the right as small as we please, for $-a < -a_0$, $b > b_0$ and some $n > n_0$. Then by taking n sufficiently large we can make the fifth term as small as we please (without affecting the smallness of the other terms). Hence $|\,\mu_j' - \mu_j'(n)\,|$ may be made as small as we please.

This establishes the more general result. The theorem enunciated at the beginning of the section follows as a corollary. In fact, if $\mu_j'(n)$ tends to a limit μ_j', then the subsequence $F_{n'}$ can always be selected and tends to a distribution function F with the moments μ_j'. All we have to prove is that if the μ_j' are such that they uniquely determine F, the sequence F_n itself converges to F.

Suppose that there exists a point of continuity x_0 such that $F_n(x_0)$ does not converge to $F(x_0)$. Then a subsequence $F_{n'}(x)$ can be selected which converges to some other value at x_0. But from this $F_{n'}$ we can select a subsequence $F_{n''}$ converging say to $F_\nu(x)$, having the same moments as $F(x)$. Since by hypothesis these moments uniquely determine F, F_ν must be the same as F in all points of continuity, i.e.

$$\lim_{n''-} F_{n''}(x_0) = F(x_0).$$

This is impossible, for $F_{n''}(x_0)$ is a subsequence of $F_{n'}(x_0)$ which converges, but not to $F(x_0)$.

4.25. The above proof can hardly be described as easy, though it depends only on simple notions such as continuity and convergence, but the Second Limit Theorem is so important that it has seemed worth while reproducing the proof in full. Many examples of its application will occur in the sequel. The chapter may be concluded with an illustration of its use in determining the limiting forms of distributions.

Example 4.8

The discontinuous distribution whose frequency at $x = j\ (j = 0, 1, \ldots)$ is $e^{-m}\frac{m^j}{j!}$ has a characteristic function

$$\phi(t) = \exp m(e^{it} - 1),$$

and hence all cumulants equal to m.

The distribution is evidently the only one with such cumulants, for $\sum_1^\infty \kappa_j \frac{(it)^j}{j!} = m\Sigma\frac{(it)^j}{j!}$ is convergent and equals $m(e^{it} - 1)$, so that the cumulative function and the characteristic function are uniquely determined.

Now as m tends to infinity the frequency at x_j, $e^{-m}\frac{m^j}{j!}$, tends to zero and thus the distribution does not tend to a limit. This is consistent with the behaviour of the cumulants, which increase without limit.

Suppose, however, we express the distribution in standard measure. Then

$$\frac{\kappa_r}{\kappa_2^{\frac{r}{2}}} = \frac{m}{m^{\frac{r}{2}}}$$

Hence as $m \to \infty$ all cumulants higher than the second tend to zero, and hence the cumulants of the distribution tend to those of the normal distribution

$$dF = \frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}(x-\sqrt{m})^2}\, dx, \qquad -\infty \leqslant x \leqslant \infty$$

with the mean $m^{\frac{1}{2}}$.

Now we know that this distribution is completely determined by its moments (Example 4.7). We also know that the cumulants determine the moments and vice-versa, so that if the cumulants of the discontinuous distribution tend to those of the normal distribution, the moments will tend to the moments of that distribution. Hence the Second Limit Theorem is applicable, and the discontinuous distribution does in fact tend to the normal form *when expressed in standard measure.*

NOTES AND REFERENCES

The idea of the characteristic function can be traced back as far as Laplace, but its introduction into the theory of statistics, through the theory of probability, is mainly due to Poincaré and Lévy (1925), whose book provides the most readable and complete account of the function. More recent researches are outlined by Cramér (1937). The proof of the First Limit Theorem is substantially that given by Lévy. The converse, given originally in a somewhat less general form by Lévy, was proved simultaneously by him and Cramér, the proof in **4.12** following the latter's.

The Second Limit Theorem seems to have been first proved by Markoff for the case when the limiting form is the normal distribution $dF = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx$. It was subsequently considered and extended by several writers, the general form of **4.24** being due to Fréchet and Shohat (1931), whose proof has been closely followed. Some references to prior work are given by these authors.

The problem of moments appears to have been first considered and solved by

Tchebycheff. The memoir by Stieltjes (1918—the memoir being first published in 1894) is classical. For some subsequent work see Hamburger (1920) and Carleman (1925).

Carleman, T. (1925), *Les fonctions quasi-analytiques*, Gauthier-Villars, Paris.

Cramér, H. (1937), *Random Variables and Probability Distributions*, Cambridge University Press.

Fréchet, M., and Shohat, J. (1931), " A Proof of the Generalised Second-Limit Theorem," *Trans. Am. Math. Soc.*, **33**, 533.

Hamburger, H. (1920, 1921), " Über eine Erweiterung des Stieltjesschen Momentproblems," *Math. Annalen*, **81**, 235, and **82**, 120 and 168.

Lévy, P. (1925), *Calcul des probabilités*, Gauthier-Villars, Paris.

Stieltjes, J. (1918), *Recherches sur les fractions continues*, *Œuvres*, Groningen.

EXERCISES

4.1. Show that if a frequency function $f(x)$ is symmetrical the characteristic function is an even function, i.e. $\phi(t) = \phi(-t)$, and that therefore $\phi(t)$ is real; and conversely, if $\phi(t)$ is real the frequency function, if any, is symmetrical.

4.2. Show that the function

$$\phi(t) = \left(\frac{e^{it} - 1}{it}\right)^n, \quad n \text{ a positive integer,}$$

is the characteristic function of a distribution function

$$F(x) = \frac{1}{n!}\left\{x^n - \binom{n}{1}(x-1)^n + \binom{n}{2}(x-2)^n \; . \; . \; .\right\}.$$

4.3. Show that the factorial moment-generating function $\omega(t)$ of the binomial $(q+p)^n$ is $(1+pt)^n$, and hence that

$$\mu'_{[r]} = p^r \, n^{[r]}.$$

4.4. If for a certain distribution

$$\kappa_r = ba^r,$$

a and b being positive constants, show that the distribution is discontinuous with variate-values $0, a, \ldots, ra, \ldots$ and the frequency at ra equal to $\dfrac{e^{-b}b^r}{r!}$.

4.5. Show that the function $e^{-t^{\alpha}}$ cannot be a characteristic function unless $\alpha = 2$.

4.6. Show that there is only one distribution with moments given by

$$\mu_r' = \frac{\Gamma(\nu + r)}{\Gamma(\nu)}$$

and that it is

$$dF = \frac{1}{\Gamma(\nu)} e^{-x} x^{\nu-1}\, dx \qquad 0 \leqslant x \leqslant \infty.$$

4.7. A theorem due to Weierstrass states that any function continuous in the range (a, b) can be represented by a uniformly convergent series of polynomials $\sum_{n=0}^{\infty} P_n(x)$, $P_n(x)$ being of degree n in x. Deduce that if two continuous frequency functions, f_1 and f_2, have the same moments of all orders,

$$\int_a^b (f_1 - f_2)^2 \, dx = 0,$$

and hence that the moments determine a distribution uniquely if it is continuous and of finite range.

4.8. If θ is a non-negative function of the variate x and

$$\alpha(t) = \int_{-\infty}^{\infty} \theta^t \, dF(x),$$

show that the frequency function of θ, if any, is given by

$$f(\theta) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \theta^{-t-1} \alpha(t) \, dt.$$

4.9. Show that if a characteristic function $\phi(t)$ possesses derivatives up to and including the second order, then

$$\left|\left(\frac{d\phi}{dt}\right)^2_{t=}\right| \qquad \left|\left(\frac{d^2\phi}{dt^2}\right)_{t=0}\right|$$

and generalise this result.

4.10. A theorem of Denjoy (*Comptes rendus*, 1921, **173**, 1399) states that if a function $f(x)$ defined in a range (a, b) possesses derivatives of all orders, if M_n is the maximum of $|f(n)(x)|$ in the range and if $\Sigma\frac{1}{M_n^{\frac{1}{n}}}$ is divergent, then $f(x)$ is completely determined by its value and that of its derivatives at a single point. Use this result to show that a set of moments determines a distribution uniquely if $\Sigma\frac{1}{\nu_n^{\frac{1}{n}}}$ diverges.

CHAPTER 5

STANDARD DISTRIBUTIONS—(1)

5.1. There are certain distribution and frequency functions which, for both theoretical and practical reasons, occupy a central position in statistical theory. In this and the next chapter we shall consider their properties, leaving their statistical uses to be developed and illustrated later in the book. We shall, however, indicate briefly some of the ways in which they arise, even at the expense of anticipating ideas introduced at a subsequent stage. This will not impair the logical continuity of our development and will give concreteness to a treatment which might otherwise appear somewhat abstract.

The Binomial Distribution

5.2. Suppose we have a large population of members each of which exhibits either some quality P or a complementary quality Q ($=$ not-P), for example, a population of men who are either blue-eyed or not-blue-eyed. Suppose that the proportion of individuals with quality P is p and that with quality Q is q, where of course $p + q = 1$. If we take a random sample of N members from the population we expect that on the average pN members will exhibit P and Nq will exhibit Q. We may thus array the members according to the quality as

$$N(p + q).$$

Now suppose we choose N *pairs* of individuals. There will be pairs PP, pairs PQ, pairs QP and pairs QQ. Of the Np pairs for which the first member is P there will, on the average, be a proportion p for which the second member is P and q for which it is Q. Similarly for the Nq exhibiting Q in the first member. Thus the pairs may be arrayed as

$$Np(p + q) + Nq(p + q) = N(p + q)^2.$$

Generally if we choose N sets of n the array will be $N(p + q)^n$. That is to say, the proportion of cases containing j P's and $(n - j)$ Q's will be $\binom{n}{j}p^j q^{n-j}$, the term in $p^j q^{n-j}$ in $(p + q)^n$. We are then led to consider the binomial distribution

$$f = (p + q)^n \quad . \quad . \quad . \qquad (5.1)$$

as a discontinuous frequency-distribution, the variate being the number of P's in the set of n, which may vary from n to 0. If, as is frequently more convenient, we wish to consider the variate as increasing from 0 to n, the distribution is inverted, i.e. becomes

$$f = (q + p)^n \qquad (5.2)$$

5.3. Distributions very close to the binomial form occur in practice, particularly in artificial experiments with coin-tossing or dice-throwing. Some data by Weldon are shown in Table 5.1. Weldon threw 12 dice 26,306 times and noted the values at each throw. This is equivalent to the drawing of samples of 12 from a large population. The occurrence of a 5 or a 6 on any die was regarded as the exhibition of the quality P, a "success" as we may call it.

TABLE 5.1

Frequency-distribution of 26,306 Throws of 12 Dice, the Occurrence of a 5 or 6 being counted a Success.

No. of Successes.	Observed Frequency.	Theoretical Frequency from the Binomial 26,306 $(0{\cdot}6623 + 0{\cdot}3377)^{12}$	No. of Successes.	Observed Frequency.	Theoretical Frequency from the Binomial 26,306 $(0{\cdot}6623 + 0{\cdot}3377)^{12}$
0	185	187	6	3,067	3,043
1	1,149	1,146	7	1,331	1,330
2	3,265	3,215	8	403	424
3	5,475	5,465	9	105	96
4	6,114	6,269	10 and over	18	16
5	5,194	5,115			
			TOTAL	26,306	26,306

If the dice were perfect (a condition rarely realised in practice) the proportion p of successes would be $\frac{1}{3}$; and the appropriate binomial would be, in the form (5.2), $(\frac{2}{3} + \frac{1}{3})^{12}$. In this particular case the dice were not quite perfect, the proportion of cases exhibiting a 5 or a 6 being 0·3377. Taking this as the value of p, we get the frequency function $(0{\cdot}6623 + 0{\cdot}3377)^{12}$, which when multiplied by the total frequency 26,306 gives the theoretical frequencies shown in the third column of Table 5.1. The agreement with observation is evidently fairly good.

5.4. Taking our variate to be increasing, we have, from (5.2), that the frequency at $x = j$ is $\binom{n}{j}q^{n-j}p^j$. The characteristic function of the distribution is then

$$\begin{aligned}\phi(t) &= \sum_{j=0}^{n} \binom{n}{j} q^{n-j}p^j e^{ijt} \\ &= (q + pe^{it})^n \qquad (5.3)\end{aligned}$$

We then have for the moment of order j about the origin, from (3.11),

$$\mu'_j = \frac{1}{i^j}\left[\frac{d^j}{dt^j}(q + pe^{it})^n\right]_{t=0}$$

and hence

$$\left.\begin{aligned}\mu'_1 &= np \\ \mu'_2 &= np + n(n-1)p^2\end{aligned}\right\} \qquad (5.4)$$

and so on. We find

$$\mu_2 = npq \qquad (5.5)$$

$$\mu_3 = npq(q - p) \qquad (5.6)$$

$$\mu_4 = 3n^2p^2q^2 + pqn(1 - 6pq) \qquad (5.7)$$

and hence

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{q-p}{(npq)^{\frac{1}{2}}} = \beta_1^{\frac{1}{2}} \qquad (5.8)$$

$$\gamma_2 = \frac{\mu_4 - 3\mu_2^2}{\mu_2^2} = \frac{1 - 6pq}{npq} = \beta_2 + 3. \qquad (5.9)$$

5.5. Further formulae are not often required, but when they are can be derived from some interesting recurrence relations connecting the moments of the binomial.

Writing $\theta = it$ we have, for the characteristic function referred to the mean as origin,

$$\phi(t) = e^{-np\theta}(q + pe^{\theta})^n. \qquad (5.10)$$

Differentiating with respect to θ we find

$$\sum_{j=1}^{n} \frac{\mu_j \theta^{j-1}}{(j-1)!} = -np\, e^{-np\theta}(q + pe^{\theta})^n + ne^{-np\theta}(q + pe^{\theta})^{n-1}\, pe^{\theta}$$

$$= -np \sum_{j=0}^{n} \frac{\mu_j \theta^j}{j!} + \frac{npe^{\theta}}{q + pe^{\theta}} \sum_{j=0}^{n} \frac{\mu_j \theta^j}{j!}$$

and hence, after a little re-arrangement,

$$(q + pe^{\theta})\left\{\sum_{j=1}^{n} \frac{\mu_j \theta^{j-1}}{(j-1)!}\right\} - npq(e^{\theta} - 1)\sum_{j=0}^{n} \frac{\mu_j \theta^j}{j!} = 0.$$

Identifying coefficients in θ^{r-1} we get

$$\mu_r = npq \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j - p \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_{j+1} \qquad (5.11)$$

giving the moment of order r about the mean in terms of those of lower orders.

Furthermore, writing the moment about the mean as

$$\mu_r = \sum_{j=0}^{n} (j - np)^r \binom{n}{j} q^{n-j} p^j$$

we have, differentiating with respect to p,

$$\frac{d\mu_r}{dp} = -rn\Sigma(j-np)^{r-1}\binom{n}{j}q^{n-j}p^j - \Sigma(j-np)^r\binom{n}{j}q^{n-j-1}(n-j)p^j + \Sigma(j-np)^r\binom{n}{j}q^{n-j}jp^{j-1}.$$

The first term on the right is $-rn\mu_{r-1}$. The sum of the other two will be found to be $\frac{1}{pq}\Sigma(j - np)^{r+1}\binom{n}{j}q^{n-j}p^j = \frac{1}{pq}\mu_{r+1}$. Hence we find

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right) \qquad (5.12)$$

For example, $\mu_1 = 0$, $\mu_2 = npq = np(1 - p)$ and hence

$$\begin{aligned} \mu_3 &= pq\{n - 2np\} \\ &= npq(q - p) \end{aligned}$$

as stated in (5.6).

For factorial moments the expressions assume a particularly simple form. In fact, differentiating $(q + p)^n$ r times partially with respect to p and multiplying by p^r we have

$$\begin{aligned} n^{[r]}(q + p)^{n-r}p^r &= \sum_{j=r}^{n} \binom{n}{j} j^{[r]} q^{n-j} p^j \\ &= \mu_{[r]}\,; \end{aligned}$$

and hence since

$$q + p = 1$$

$$\mu_{[r]} = n^{[r]}p^r.$$

5.6. If $p = q$ the binomial distribution is obviously symmetrical. If $p \neq q$ the distribution is skew. But in both cases it will be unimodal unless pn is small. For the frequency of the $(r + 1)$th term is greater than that of the rth so long as

$$\binom{n}{r+1} q^{n-r-1} p^{r+1} > \binom{n}{r} q^{n-r} p^{r}$$

or

$$\frac{n!}{(n-r)!r!} \frac{(n-r-1)!(r+1)!}{n!} < \frac{p}{}$$

or

$$\frac{r+1}{n-r} < \frac{p}{q},$$

which is equivalent to

$$\frac{r+1}{n+1} < p.$$

Hence the frequency increases until the point when $(r + 1) > p(n \quad 1)$ and then declines again. Some typical distributions are shown in Table 5.2.

TABLE 5.2

Terms of the Binomial Distribution 10,000 $(q + p)^{20}$ for Values of p from 0·1 to 0·5.
(Figures given to the nearest unit.)

Number of Successes.	$p = 0{\cdot}1$ $q = 0{\cdot}9$	$p = 0{\cdot}2$ $q = 0{\cdot}8$	$p = 0{\cdot}3$ $q = 0{\cdot}7$	$p = 0{\cdot}4$ $q = 0{\cdot}6$	$p = 0{\cdot}5$ $q = 0{\cdot}5$
0	1216	115	8	—	—
1	2702	576	68	5	—
2	2852	1369	278	31	2
3	1901	2054	716	123	11
4	898	2182	1304	350	46
5	319	1746	1789	746	148
6	89	1091	1916	1244	370
7	20	545	1643	1659	739
8	4	222	1144	1797	1201
9	1	74	654	1597	1602
10	—	20	308	1171	1762
11	—	5	120	710	1602
12	—	1	39	355	1201
13	—	—	10	146	739
14	—	—	2	49	370
15	—	—	—	13	148
16	—	—	—	3	46
17	—	—	—	—	11
18	—	—	—	—	2
19	—	—	—	—	—
20	—	—	—	—	—

5.7. The ordinates of the binomial are most directly calculated from the formula $\binom{n}{j} q^{n-j} p^{j}$; for low values of n the calculation is straightforward and for high values assistance can be derived from tables of log n! The calculation of the distribution function,

which is equivalent to the summation of terms of the binomial, is tedious to perform directly, but use may be made of the tables of the incomplete B-function. We have, for Taylor's series with the integral form of remainder

$$f(a+h)=\sum_{j=0}^{r-1}\frac{h^j}{j!}f^{[j]}(a)+\int_0^1\frac{h^r(1-t)^{r-1}}{(r-1)!}f^{(r)}(a+th)\,dt. \tag{5.13}$$

Putting $a=q$, $h=p$ and $f(a+h)=(q+p)^n$ we have

$$(q+p)^n=\sum_{j=0}^{r-1}\binom{n}{j}q^{n-j}p^j+R_r,$$

where R_r is the remainder after r terms and equals

$$R_r=\int_0^1\frac{p^r(1-t)^{r-1}}{(r-1)!}\frac{n!}{(n-r)!}(q+pt)^{n-r}\,dt \tag{5.14}$$

In (5.14) put $t=1-\dfrac{x}{p}$. We find

$$\begin{aligned}R_r&=\frac{n!}{(r-1)!(n-r)!}\int_0^p x^{r-1}\,(1-x)\qquad dx\\&=\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}B_p(r,n-r+1)\\&=\frac{B_p(r,n-r+1)}{B(r,n-r+1)}\\&=I_p(r,n-r+1)\end{aligned} \tag{5.15}$$

iu the usual notation. This is also equal to

$$1-I_q(n-r+1,r)\ . \tag{5.16}$$

by a well-known property of the B-function.

The remainder after $r-1$ terms is, similarly, $I_p(r-1,n-r+2)$, and hence the rth term is

$$\begin{aligned}&-I_p(r,n-r+1)+I_p(r-1,n-r+2)\\&=I_q(n-r+1,r)-I_q(n-r+2,r-1).\end{aligned} \tag{5.17}$$

Example 5.1

When $n=20$, $r=11$, $p=0{\cdot}4$ we have for the remainder after 11 terms $I_{0\cdot4}(11,10)$ which from the tables is found to be 0·127,521,2. The value given by summing the last six terms in the appropriate column of Table 5.2 is 0·1276, the error in the last place being due to rounding up. The remainder after 12 terms is $I_{0\cdot4}(12,9)=0{\cdot}056{,}526{,}4$. The 11th term (10 " successes ") is then the difference of these two remainders $=0{\cdot}0710$, as shown in Table 5.2 for the frequency per 10,000 of 11 successes.

The Poisson Distribution

5.8. Cases sometimes occur in which the proportion p of " successes " in the population is very small. We may suppose our number n large enough to render np itself appreci-

able though p is small ; and we are thus led to consider the limiting form of the binomial (5.2) as $p \to 0$ subject to the condition that np remains finite, and equal to λ, say.

Under these conditions the term

$$\binom{n}{r}p^r q^{n-r} = \frac{n!}{(n-r)!r!}\frac{\lambda^r}{n^r}\left(1-\frac{\lambda}{n}\right)^{n-r}$$
$$\sim \frac{\sqrt{(2\pi)}e^{-n}n^{n+\frac{1}{2}}}{\sqrt{(2\pi)}(n-r)^{n-r+\frac{1}{2}}e^{-n+r}n^r}\cdot\frac{\lambda^r}{r!}e^{-\lambda}$$
$$\sim \frac{1}{\left(1-\frac{r}{n}\right)^n e^r}\cdot\frac{\lambda^r}{r!}e^{-\lambda} \sim \frac{\lambda^r}{r!}e^{-\lambda}.$$

Thus the terms of the binomial become the successive terms

$$e^{-\lambda}\left(1, \frac{\lambda}{1!}, \ldots \frac{\lambda^j}{j!}, \ldots\right) \quad . \qquad (5.18)$$

This is called the Poisson distribution, having been given first by Poisson in 1837. It has since been discovered independently by several other writers.

From the point of view of characteristic functions we have

$$\phi(t) = \lim\, (q+pe^{it})^n$$
$$= \lim \left\{1+\frac{\lambda}{n}(e^{it}-1)\right\}^n$$
$$= \exp \lambda(e^{it}-1) \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5.19)$$

which is readily verified to result in the distribution (5.18).

Thus

$$\psi(t) = \lambda(e^{it}-1) = \lambda\Sigma\frac{(it)^j}{j!}$$

and hence all cumulants of the Poisson distribution are equal to λ. We thus find

$$\left.\begin{aligned}\mu_1 &= \lambda\\ \mu_2 &= \lambda\\ \mu_3 &= \lambda\\ \mu_4 &= \lambda+3\lambda^2\end{aligned}\right\} \qquad (5.20)$$

If we let $n \to \infty$ in (5.11) and (5.12) we find

$$\mu_r = \lambda\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_j \qquad (5.21)$$

and

$$\mu_{r+1} = r\lambda\mu_{r-1} + \frac{\lambda\, d\mu_r}{d\lambda}\, . \qquad (5.22)$$

5.9. Tables of the function $e^{-\lambda}\frac{\lambda^r}{r!}$ for various values of λ and r are given in *Tables for Statisticians and Biometricians*, Part I. The frequency polygons are very skew, almost J-shaped for low values of λ, but become nearer to unimodal symmetry as λ increases. A comparison of the successive terms $\frac{\lambda^r}{r!}$ and $\frac{\lambda^{r+1}}{(r+1)!}$ shows that the frequency increases up to the point for which $r+1 < \lambda$ and then decreases again.

The summation of r terms in the Poisson distribution may be carried out in a manner similar to that of **5.7**. The remainder after r terms of the distribution is found to be, from (5.13)

$$\begin{aligned} R_r &= \frac{\lambda^r}{\Gamma(r)}\int^1 e^{\lambda(1-t)}(1-t)^{r-1}\,dt \\ &= \frac{\Gamma_\lambda(r)}{\Gamma(r)} \\ &= I\left(\frac{\lambda}{\sqrt{r}}, r-1\right), \end{aligned} \qquad (5.23)$$

in the notation of Pearson's tables of the Incomplete Γ-function. The argument used in these tables is a difficult one to work with in the present case and, though formula (5.23) may be used for summing a number of terms in the Poisson distribution, it is easier to calculate $e^{-\lambda}\dfrac{\lambda^r}{r!}$ directly rather than to use an analogous expression to (5.17) in the form

$$r\text{th term} = I\left(\frac{\lambda}{\sqrt{(r-1)}}, r-2\right) - I\left(\frac{\lambda}{\sqrt{r}}, r-1\right).$$

5.10. We now consider a generalisation of the binomial and the Poisson distributions. In **5.2** our approach was based on the drawing of sets of n from the same population. Suppose, however, we draw them from n different populations with proportions

$$(p_1, q_1)\ (p_2, q_2)\ \ldots\ (p_n, q_n).$$

Then our proportional frequencies will be arrayed by the form

$$(p_1+q_1)(p_2+q_2)\ldots(p_n+q_n) = \prod_{j=1}^{n}(p_j+q_j) \qquad (5.24)$$

which of course reduces to the binomial if all the p's are equal.

The characteristic function of this distribution is

$$\phi(t) = \Pi(q_j + p_j e^{it})$$

from which we have

$$\begin{aligned} \psi(t) &= \Sigma \log(q_j + p_j e^{it}) \\ &= \Sigma \log\left\{1 + p_j it + p_j\frac{(it)^2}{2!} + \ldots\right\} \\ &= (it)\Sigma p_j + \frac{(it)^2}{2}\Sigma(p_j - p_j^2) +, \text{ etc.,} \end{aligned}$$

giving

$$\left.\begin{aligned} \mu_1' &= \Sigma p_j \\ \kappa_2 = \mu_2 &= \Sigma p_j q_j \end{aligned}\right\}. \qquad (5.25)$$

Writing now $\bar{p}$ for the mean of the p's in the different populations, we have

$$\begin{aligned} \mu_1' &= n\bar{p} \\ \mu_2 &= \Sigma pq = \Sigma p - \Sigma p^2 \\ &= \Sigma p - \frac{1}{n}(\Sigma p)^2 - \left\{\Sigma p^2 - \frac{1}{n}(\Sigma p)^2\right\} \\ &= n\bar{p} - n\bar{p}^2 - n \operatorname{var} p \end{aligned}$$

(where var p is written for the variance of p)

$$= n\bar{p}\bar{q} - n \operatorname{var} p \qquad (5.26)$$

A comparison of these results with those for the binomial shows that the variance of the distribution (5.24) is *less* than that of the binomial with the same average p by an amount equal to n times the variance of p.

Similarly we see that, for the Poisson distribution in such a case

$$\mu_1' = \bar{\lambda}, \text{ the mean of the } \lambda\text{'s} \quad . \quad . \quad . \qquad (5.27)$$

and

$$\mu_2 = \bar{\lambda} - \frac{1}{n} \operatorname{var}(np)$$
$$= \bar{\lambda} \text{ to order } n^{-1}.$$

The Poisson form thus holds for (5.24) notwithstanding the inequality of the p's, provided that the variance of λ is small compared with n, which will be so if all the p's are small.

5.11. Consider now the case when successive sets of n are drawn from different populations characterised by $p_1, p_2 \ldots p_k$. In the previous case we supposed any set of n obtained by taking one from each of n populations.

We now suppose that any set is drawn from one population only, but that different sets come from different populations. Our array of frequencies will now be

$$\frac{1}{k}\sum_{j=1}^{k}(q_j + p_j)^n \qquad (5.28)$$

and evidently the moments of this array are the sums of the moments of the $(q + p)^n$, that is to say, from (5.4),

$$\mu_1' = \frac{1}{k}\Sigma np$$

$$\mu_2 = \frac{1}{k}\{\Sigma np + \Sigma n(n - 1)p^2\}.$$

Writing $\bar{p}$ for the mean of the p's as before, we have

$$\mu_1' = n\bar{p}$$
$$\mu_2 = n\bar{p} + \frac{1}{k}\Sigma n(n - 1)p^2 - n^2\bar{p}^2$$
$$= n\bar{p}\bar{q} + \frac{1}{k}\Sigma n(n - 1)p^2 - n(n - 1)\bar{p}^2$$
$$= n\bar{p}\bar{q} + n(n - 1)\left\{\frac{1}{k}\Sigma p^2 - \bar{p}^2\right\}$$
$$= n\bar{p}\bar{q} + n(n - 1) \operatorname{var} p. \quad . \quad . \quad . \quad . \quad . \qquad (5.29)$$

In this case the variance is *greater* than what it would be if the distribution were of the ordinary binomial type by an amount $n(n - 1) \operatorname{var} p$.

For the Poisson distribution we have, on taking limits,

$$\left.\begin{aligned}\mu_1' &= \bar{\lambda}\\ \mu_2 &= \bar{\lambda} + \operatorname{var} \lambda\end{aligned}\right\} \qquad (5.30)$$

and here also the variance of the distribution is affected.

5.12. The results of the two preceding sections enable us to discuss the occurrence of the binomial and the Poisson distributions in practice. An example has already been

given in Table 5.1 of a distribution conforming to the simple binomial type. It is not easy to find material compiled outside the laboratory which does so.

For example, suppose we regard the possession of blue eyes as a success, and take a number of samples of n from the population of the United Kingdom in different localities. We should probably find that the proportions in these samples did not conform to the simple binomial form. The variance npq calculated from the known n and observed p would probably turn out to be too small. If so we should conclude from (5.26) that the proportion p varied from place to place in the population, the deficiency in the variance of the proportions observed being due to the variance of p itself in the sections of the population from which the samples were chosen. We are assuming for the time being that these differences are not explicable on the basis of sampling fluctuation alone; but a full discussion will have to wait until later chapters.

5.13. The same effect is found in distributions which at first sight might be expected to be of the Poisson type. For example, suicide is a rare event and it might be supposed that if we took a series of large samples, say the population of the United Kingdom in successive years, the frequencies of suicides would follow the Poisson distribution. This, however, is not necessarily so, for all members of the population are not equally exposed to risk and the temptation to suicide may vary from year to year, e.g. being greater in years of trade depression. This inequality of risk is typical of one field in which the Poisson distribution has been freely applied, namely, industrial accidents. Table 5.3 shows, in the second column, the frequency of accidents occurring to women working on the manufacture of shells. The Poisson frequencies shown in the third column provide a very poor fit. The reason is that the liability of individuals to accident varies.

TABLE 5.3

Accidents to 647 women working on H.E. shells in 5 weeks.

(Greenwood and Yule (1920), *J. Roy. Statist. Soc.*, **83**, 255.)

Number of Accidents.	Observed Frequency.	Poisson Distribution with same Mean.	Distribution given by (5.33).
0	447	406	442
1	132	189	140
2	42	45	45
3	21	7	14
4	3	1	5
5	2	0·1	2
TOTALS	647	648	648

As a working hypothesis (cf. Greenwood and Yule, 1920) suppose that the population is composed of individuals with different degrees of accident proneness, represented by different values of λ in a Poisson distribution; and suppose that in the population the distribution of λ is given by

$$dF = \frac{c^p}{\Gamma(p)} e^{-c\lambda}\, \lambda^{p-1}\, d\lambda \qquad 0 \leqslant \lambda \leqslant \infty . \qquad (5.31)$$

There are theoretical reasons justifying this supposition.

The frequency of j successes is then

$$\int_0^{\infty} \frac{c^p}{\Gamma(p)} e^{-c\lambda}\lambda^{p-1}e^{-\lambda}\frac{\lambda^j}{j!}\,d\lambda$$

or the coefficient of t^r in

$$\frac{c^p}{\Gamma(p)}\int_0^{\infty} e^{-c\lambda}\,\lambda^{p-1}\,e^{-\lambda+\lambda t}\,d\lambda,$$

which, on the substitution of $(c + 1 - t)\lambda = u$, becomes

$$\frac{c^p}{(c+1-t)^p} = \left(\frac{c}{c+1}\right)^p\left(1 - \frac{t}{c+1}\right)^{-p} \tag{5.32}$$

The frequency of 0, 1, 2, . . . successes is therefore

$$\left(\frac{c}{c+1}\right)^p\left\{1,\ \frac{p}{c+1},\ \frac{p(p+1)}{2!(c+1)^2}\ \ldots\right\} \tag{5.33}$$

The mean is thus

$$\begin{aligned}\mu_1' &= \left(\frac{c}{c+1}\right)^p\left\{\frac{p}{c+1} + \frac{p(p+1)}{(c+1)^2} + \ldots\right\}\\ &= \left(\frac{c}{c+1}\right)^p.\frac{p}{c+1}.\left(\frac{c+1}{c}\right)^{p+1}\\ &= \frac{p}{c}\end{aligned} \tag{5.34}$$

Similarly $\qquad \mu_2' = \dfrac{p}{c}\left(\dfrac{p+c+1}{c}\right)$, so that

$$\mu_2 = \frac{p(c+1)}{c^2} = \frac{p}{c} + \frac{p}{c^2} \tag{5.35}$$

If we now put the observed mean and variance of Table 5.3 equal to the values of (5.34) and (5.35) we have two equations which can be solved for p and c. The distribution (5.33) can then be found. The frequencies are given in the fourth column of Table 5.3 and evidently give a much better agreement with the facts.

5.14. The interesting feature of the distribution (5.32) is that it is a binomial *with negative index*. In the approach adopted in **5.2** the index is necessarily positive; but it is often found that observational materials are represented by negatively indexed binomials. Yule (1910) * has given an illustration of this effect which does not depend on any arbitrary assumption about distributions such as that embodied in (5.31). Suppose, in fact, that we have a population subjected to recurring attacks of a disease, that r attacks are fatal and that on the average one attack is fatal to a proportion p of individuals at risk, the actual numbers succumbing varying as if the population were chosen at random from a larger population in which the proportion of survivors is p. Consider the proportion of individuals surviving 0, 1, . . . attacks at the nth exposure. Evidently this is the proportion of successes in samples of n when the chance of success is p, i.e. $(q + p)^n$. The proportion of survivors at the end of n exposures will be the sum of the first r terms in this series.

* *J. Roy. Statist. Soc.*, **73**, 26.

The proportion of survivors at the end of $(n-1)$ exposures will be the sum of the first r terms in $(q+p)^{n-1}$. Consequently the proportion dying during the nth exposure is the difference,

$$\begin{aligned}&\sum_{j=0}^{r-1}\binom{n}{j}q^{n-j}p^j - \sum_{j=0}^{r-1}\binom{n-1}{j}q^{n-j-1}p^j\\ &= \sum_{j=0}^{r-1}\left\{\binom{n-1}{j}q^{n-j}p^j + \binom{n-1}{j-1}q^{n-j}p^j - \binom{n-1}{j}q^{n-j-1}p^j\right\}\\ &= \sum_{j=0}^{r-1}\left\{\binom{n-1}{j-1}q^{n-j}p^j - \binom{n-1}{j}q^{n-j-1}p^{j+1}\right\}\\ &= \binom{n-1}{r-1}q^{n-r}p^r.\end{aligned}$$

Thus, since death does not commence till the rth exposure, for the values of n from r onwards we have the proportion of deaths

$$p^r\{1, \quad rq, \quad \frac{r(r+1)}{2!}q^2, + \ldots \qquad (5.36)$$

i.e. successive terms in $p^r(1-q)^{-r}$, a binomial with negative index. A law of this kind has been found to operate in experiments on the killing of bacteria by disinfectants.

The Hypergeometric Distribution

5.15. Consider now the generalisation of the approach of **5.2** when samples of n are drawn from a population of N individuals, where N is not necessarily large. If we take a sample which contains r P's and $n-r$ Q's, it can arise in

$$\begin{aligned}&\binom{n}{r}\frac{Np(Np-1)\ldots(Np-r+1)Nq(Nq-1)\ldots(Nq-n+r-1)}{N(N-1)\ldots(N-n+1)}\\ &\qquad = \binom{n}{r}\frac{(Np)^{[r]}(Nq)^{[n-r]}}{N^{[n]}} \qquad . \quad . \quad . \quad . \quad . \quad (5.37)\end{aligned}$$

ways. For there are $\binom{N}{n}$ ways of selecting the sample, and the r P's can be chosen in $\binom{Np}{r}$ ways, and the $n-r$ Q's in $\binom{Nq}{n-r}$ ways, the expression given in (5.37) being equal to $\binom{Np}{r}.\binom{Nq}{n-r}\Big/\binom{N}{n}$

Hence we are led to consider the discontinuous distribution

$$f = \frac{1}{N^{[n]}}\sum_{j=0}^{n}\left\{\binom{n}{j}(Np)^{[j]}(Nq)^{[n-j]}\right. \qquad (5.38)$$

a form in which the analogy with the binomial (5.1) is evident. As $N \rightarrow \infty$ the form (5.38) approaches the binomial.

The series

$$f(x) = \frac{1}{N^{[n]}}\sum_{j=0}^{n}\left\{\binom{n}{j}(Np)^{[j]}(Nq)^{[n-j]}x^j\right. \qquad (5.39)$$

is equal to

$$\frac{(Nq)^{[n]}}{N^{[n]}} \quad \frac{(Np)^{[j]}n^{[j]}}{(Nq-n+j)^{[j]}}\frac{x^j}{j!}$$

that is to say, to the hypergeometric function

$$\frac{(Nq)^{[n]}}{N^{[n]}} \quad F(\alpha, \beta\,;\, \gamma, x)$$

if

$$\alpha = -n, \quad \beta = -Np, \quad \gamma = Nq - n + 1. \qquad (5.40)$$

The distribution (5.38) is therefore called hypergeometric. We have

$$F(\alpha, \beta\,;\, \gamma, x) \quad 1 + \frac{\alpha\beta}{\gamma}\frac{x}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{x^2}{2!} +$$

and it is well known that this function satisfies the differential equation

$$x(1-x)\frac{d^2F}{dx^2} + \{\gamma - (\alpha+\beta+1)x\}\frac{dF}{dx} - \alpha\beta F = 0, \qquad (5.41)$$

a fact which may be readily verified from the equation itself.

If in (5.39) we put $x = e^\theta$ ($\theta = it$) we evidently have the characteristic function of the distribution. On making this substitution in (5.41) we find, after some reduction and replacement of the values of α, β, γ by those of (5.40),

$$(1-e^\theta)\left\{\frac{d^2\phi}{d\theta^2} - (n+Np)\frac{d\phi}{d\theta} + nNp\phi\right\} - Nnp\phi + N\frac{d\phi}{d\theta} = 0\,. \qquad (5.42)$$

Since $\phi = \Sigma\frac{\mu_j'\theta^j}{j!}$ we find, from the coefficient of θ^0 in this expression,

$$-Nnp + N\mu_1' = 0$$

$$\mu_1' = np \qquad (5.43)$$

the same result as for the binomial. The mean of the hypergeometric series is independent of N.

Taking now the distribution about its mean, and hence substituting $e^{np\theta}\phi$ for ϕ in (5.42), we find

$$(1-e^\theta)\left[\frac{d^2\phi}{d\theta^2} + \frac{d\phi}{d\theta}\{n(p-q) - Np\} + (N-n)pqn\phi\right] + N\frac{d\phi}{d\theta} = 0 \qquad (5.44)$$

whence, identifying coefficients in θ, θ^2, θ^3 we find

$$\mu_2 = \frac{npq(N-n)}{N-1} \qquad (5.45)$$

$$\mu_3 = \frac{npq(q-p)(N-n)(N-2n)}{(N-1)(N-2)} \qquad (5.46)$$

$$\mu_4 = \frac{npq(N-n)}{(N-1)(N-2)(N-3)}[N(N+1) - 6n(N-n) + 3pq\{N^2(n-2) - Nn^2 + 6n(N-n)\}] \qquad (5.47)$$

and generally, if E denotes the operation of raising the order of a moment by unity, i.e. $E\mu_r = \mu_{r+1}$, we have

$$N\mu_{r+1} = \{(1+E)^r - E^r\}[\mu_2 - \{Np + n(q-p)\}\mu_1 + \{npq(N-n)\mu_0\}] \qquad (5.48)$$

As we expect, when $N \rightarrow \infty$ these values tend to those of the binomial.

5.16. An example of the occurrence of the hypergeometric series in practice is given in Table 5.4, giving the frequency of occurrence of cards of a certain suit in hands of whist. Here N is the number of cards in the pack, 52, and $n = 13$, $p = \frac{1}{4}$. The appropriate series is thus

$$\frac{1}{52^{[13]}}\Sigma\binom{13}{j}13^{[j]}39^{[n-j]}$$

giving the frequencies shown in the third column. This agreement appears to be reasonably good.

TABLE 5.4

Distribution of 3400 First Hands at Whist according to Number of Trumps in the Hand.
(K. Pearson, 1924, *Biometrika*, **16**, 172.)

Number of Cards in the Hand.	Observed Frequency.	Frequency of Hypergeometric Distribution.	Number of Cards in the Hand.	Observed Frequency.	Frequency of Hypergeometric Distribution.
0	35	43·5	5	444	424·0
1	280	272·2	6	115	141·3
2	696	700·0	7	21	30·0
3	937	973·5	8	11	4·0
4	851	811·3	9 and over	0	0·2
			TOTALS	3400	3400

5.17. The calculation of frequencies and the summing of series of frequencies is not so simple a matter as for the binomial, but the incomplete B-function may be used to give a fairly good approximation. The method consists of fitting a B-curve of type

$$dF = \frac{1}{B(p,q)}\,(1-x)^{p-1}\,x^{q-1}\,dx \qquad 0 \leqslant x \leqslant 1$$

to the distribution and obtaining areas of that curve from the B-tables. Details of the method and an example are given in the preface to the Tables of the Incomplete B-function.

The Normal Distribution

5.18. We have already noted in Examples 4.6 and 4.8 that the binomial distribution and the Poisson distribution both tend, when expressed in standard measure, to the form

$$dF = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx \qquad -\infty \leqslant x \leqslant \infty \quad . \quad (5.49)$$

The slightly more general form

$$dF = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu_1')^2}\,dx \qquad -\infty \leqslant x \leqslant \infty \qquad (5.50)$$

is known as the normal distribution. It is the most important theoretical distribution in statistics. The expression (5.49) is of course the normal distribution in standard measure.

We have for the characteristic function of (5.50)

$$\phi(t) = e^{-\frac{t^2}{2}\sigma^2} e^{it\mu'_1} \quad . \quad . \qquad . \ (5.51)$$

giving

$$\mu_{2r} = \frac{(2r)!}{2^r r!}\sigma^{2r} \qquad . \ (5.52)$$

$$\mu_{2r+1} = 0$$

and

$$\psi(t) = \frac{-t^2\sigma^2}{2} \qquad . \ (5.53)$$

so that

$$\left.\begin{aligned} \kappa_2 &= \mu_2 = \sigma^2 \\ \kappa_r &= 0, \qquad r > 2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \ (5.54)$$

We also have $\beta_2 = 3$, $\gamma_2 = 0$, which accounts for the standard adopted for mesokurtosis in **3.32.**

5.19. The distribution function of the normal distribution expressed in standard measure is the integral

$$F(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, dx. \quad . \quad . \quad . \quad . \quad . \quad . \ (5.55)$$

The integrand may be expanded and we have

$$\begin{aligned} F(x) &= \frac{1}{2} + \frac{1}{\sqrt{(2\pi)}} \int_0^x e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{2} + \frac{1}{\sqrt{(2\pi)}} \int_0^x \left\{1 - \frac{x^2}{2} + \frac{1}{2!}\left(\frac{x^2}{2}\right)^2 - \ . \ . \ . \right\} dx \\ &= \frac{1}{2} + \frac{1}{\sqrt{(2\pi)}} \left(x - \frac{x^2}{2.3} + \frac{1}{2!}\,\frac{x^5}{2^2.5} - \ . \ . \ . \text{ etc.}\right) \end{aligned} \qquad (5.56)$$

This converges too slowly to be of use for other than small values of x.

If x is large an asymptotic series may be employed. We have

$$\begin{aligned} 1 - F(x) &= \frac{1}{\sqrt{(2\pi)}} \int_x^{\infty} e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{\sqrt{(2\pi)}} \int_x^{\infty} \frac{e^{-\frac{x^2}{2}} x\, dx}{x} \\ &= \frac{1}{\sqrt{(2\pi)}} \left[-\frac{e^{-\frac{x^2}{2}}}{x} \right]_x^{\infty} - \frac{1}{\sqrt{(2\pi)}} \int_x^{\infty} \frac{e^{-\frac{x^2}{2}}}{x^3}\, dx \end{aligned}$$

and on repeating the partial integration,

$$= \frac{e^{-\frac{x^2}{2}}}{\sqrt{(2\pi)}x} \left\{1 - \frac{1}{x^2} + \frac{3}{x^4} - \frac{3.5}{x^6} \ \ldots \right\} + R_n \qquad (5.57)$$

where R_n is less in absolute value than the last term taken into account.

The most useful formula, however, is a continued fraction due to Laplace. Put

$$\alpha(t) = \frac{1}{1-t}\left\{1 - \frac{1}{x^2(1-t)^2} + \frac{3}{x^4(1-t)^4} \ . \ . \ . \right\}$$

so that $\alpha(0)$ is the expression in curly brackets in (5.57). We then have

$$\frac{1}{x^2}\frac{\partial\alpha}{\partial t} = \frac{1}{x^2(1-t)^2} - \frac{3}{x^4(1-t)^4} + \ldots$$
$$= -\{(1-t)\,\alpha(t) - 1\}$$

Hence if $\alpha(t) = \Sigma y_j t^j$ we have, identifying coefficients in t^r,

$$\frac{1}{x^2}(r+1)y_{r+1} + y_r - y_{r-1} = 0. \qquad (5.58)$$

Hence
$$\frac{y_r}{y_{r-1}} = \frac{1}{1 + \dfrac{r+1}{x^2}\dfrac{y_{r+1}}{y_r}}$$

Thus
$$\frac{y_1}{y_0} = \frac{x}{x + \dfrac{2}{x}\dfrac{y_1}{y_2}} \quad \frac{x}{x+}\;\frac{2}{x+}\;\frac{3}{x^2}\frac{y_4}{y_3}$$
$$\frac{x}{x+}\;\frac{2}{x+}\;x+ \;\ldots\; \frac{n}{x+}$$

Now when $t = 0$, α reduces to y_0, and we also see from (5.58) that as $x \to \infty, y_1 = y_0$. Hence we have, from (5.57),

$$1 - F(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{(2\pi)}\,x\,\{x + x + x + \ldots}$$
$$\frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{x^2}{2}}\left\{\frac{1}{x+}\;\frac{1}{x+}\;\frac{2}{x+}\ldots\right. \qquad (5.59)$$

The continued fraction thus gives the ratio of the frequency of the normal distribution to the right of the point x (the "tail area") to the ordinate of the frequency-distribution at that point.

This expression was in fact used by Sheppard (1939, posthumous) in calculating his superb tables of the normal function. These tables give, among other things :—

(*a*) The ratio of the tail area to the bounding ordinate, to 12 places of decimals, for intervals 0·01.

(*b*) The same to 24 places for intervals 0·1.

(*c*) The negative natural logarithm of the tail area, to sixteen places, by intervals 0·1.

Tables which are sufficient for all ordinary purposes will be found in *Tables for Statisticians and Biometricians*, Parts I and II. At the end of this volume we give some tables which will suffice to illustrate the theory and examples given herein.

5.20. The shape of the normal curve

$$y = \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{x^2}{2}}$$

is illustrated in Fig. 5.1. It is symmetrical and unlimited in range, falling off to zero very rapidly as the variate increases. There are points of inflection at unit distance on either side of the mean.

For the mean deviation we have

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|x|\,e^{-\frac{x^2}{2}}\,dx = \sqrt{\frac{2}{\pi}}\int_{0}^{\infty} xe^{-\frac{x^2}{2}}\,dx$$

$$= \sqrt{\frac{2}{\pi}} = 0{\cdot}79788 \qquad . \quad (5.60)$$

The variance is of course unity, because the distribution is expressed in standard measure. The quartiles are distant 0·674,489,75 from the mean, as may be found from the Tables.

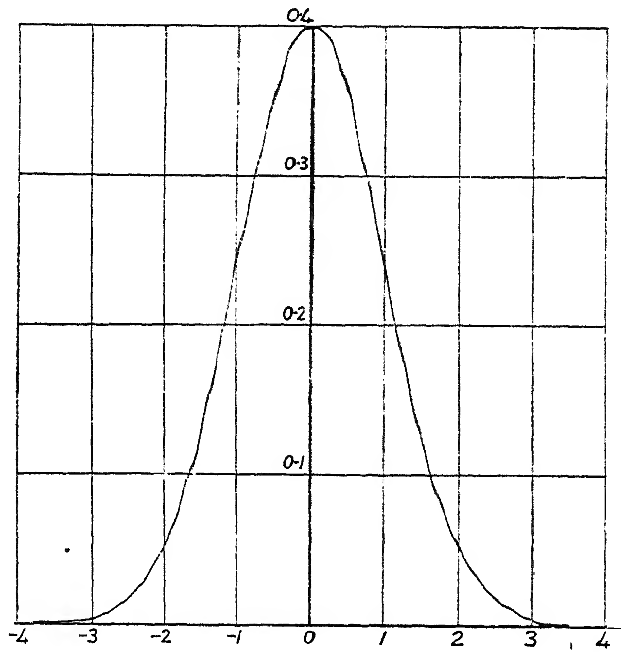


FIG. 5.1.—The Normal Curve $y = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$.

5.21. As an illustration of the occurrence in practice of a distribution which is very close to the normal, the height data of Table 1.7 may be taken. Table 5.5 shows the actual frequencies and those given by the normal curve with the same mean and standard deviation (67·46 and 2·56 inches respectively).

The correspondence is evidently fairly good. It must, however, be noted that whereas the theoretical distribution has infinite range, the practical distribution has not, since it is impossible to have a negative height. In this particular case the relative frequency of the normal distribution outside the range 57–77 inches is so small that the point is unimportant; but when distributions of finite range are represented by those of infinite range it is as well to remember that the fit near the tails may not be very close.

5.22. The normal distribution has had a curious history. It was first discovered by De Moivre in 1753 as the limiting form of the binomial, but was apparently forgotten and rediscovered later in the eighteenth century by workers engaged in investigating the theory of probability and the theory of errors. The discovery that errors of observation

TABLE 5.5

Frequency-Distribution of 8585 Men according to Height (Table 1.7) compared with Theoretical Frequencies of a Normal Distribution with the Same Mean and Variance.

Height (inches).	Observed Frequency.	Theoretical Frequency.	Height (inches).	Observed Frequency.	Theoretical Frequency.
57–	2	1	68–	1230	1234
58–	4	3	69–	1063	989
59–	14	11	70–	646	682
60–	41	33	71–	392	405
61–	83	88	72–	202	207
62–	169	200	73–	79	91
63–	394	395	74–	32	34
64–	669	669	75–	16	11
65–	990	976	76–	5	3
66–	1223	1227	77–	2	1
67–	1329	1326			
			TOTALS	8585	8586

ought, on certain plausible hypotheses, to be distributed normally led to a general belief that they *were* so distributed. The belief extended itself to distributions such as those of height, in which the variate-value of an individual may be regarded as the cumulation of a large number of small effects. Vestiges of this dogma are still found in textbooks.

It was found in the latter half of the nineteenth century that the frequency-distributions occurring in practice are rarely of the normal type and it seemed that the normal distribution was due to be discarded as a representation of natural phenomena. But as the importance of the distribution declined in the observational sphere it grew in the theoretical, particularly in the theory of sampling. It is in fact found that many of the distributions arising in that theory are either normal or sufficiently close to normality to permit satisfactory approximations by the use of the normal distribution. Furthermore, by a fortunate accident (if one may speak of accidents in mathematics) it happens that the analytic form of the normal distribution is particularly well adapted to the requirements of sampling theory. For these and other reasons which will be amply illustrated in the sequel, the normal distribution is pre-eminent among the distributions of statistical theory.

5.23. Since the normal distribution may be considered as the limit of the binomial it is natural to inquire into the limiting forms, if any, of the hypergeometric distribution. From (5.38) we see that the difference between two successive terms in the distribution is

$$\frac{1}{N^{[n]}}\frac{n!}{r!(n-r-1)!}(Np)^{[r]}(Nq)^{[n-r-1]}\left\{\frac{Np-r}{r+1}-\frac{Nq-n+r+1}{n-r}\right\}$$

$$=\frac{1}{N^{[n]}}\frac{n!}{r!(n-r)!}(Np)^{[r]}(Nq)^{[n-r]}\left\{\frac{Nnp-Nq+n-1-r(N+2)}{(r+1)(Nq-n-r)}\right\}$$

The ratio of this difference to the $(r+1)$th term is then

$$\frac{\Delta y_r}{y_r}=\frac{A+Br}{C+Dr+Er^2},$$

where the quantities $A \ldots E$ are constants. In the limit when the distribution is expressed in standard measure, Δy_r is the increment when r increases by a small quantity, and we are thus led to consider the differential equation defining a frequency function

$$\frac{df}{f} = \frac{A + Bx}{C + Dx + Ex^2}\,dx. \qquad (5.61)$$

This is the equation of a family of functions—the Pearson distributions—which will be considered from a slightly different standpoint in the next chapter.

The Bivariate Binomial Distribution

5.24. In generalisation of the results of **5.2**, consider the drawing of samples of n from a population the individuals of which may or may not have two attributes, P and not-P $(= Q)$ and R and not-R $(= S)$. Suppose that the proportions of the individuals with attributes PR, QR, PS and QS are a, b, c and d respectively, where $a + b + c + d = 1$. In exactly the same way as for the binomial case it is seen that the proportion of samples with i PR's j QR's, k PS's and l QS's is $\dfrac{n!}{i!j!k!l!}a^i b^j c^k d^l$ and the distribution of samples is given by the multinomial form

$$f = (a + b + c + d)^n. \qquad (5.62)$$

The distribution given by this form is bivariate, one variate being the number of P's and the other the number of R's. The characteristic function of the distribution is

$$\phi = (ae^{it_1+it_2} + be^{it_2} + ce^{it_1} + d)^n. \qquad (5.63)$$

We have then

$$\frac{1}{n}\log\phi = \log\Big\{a + b + c + d + ai(t_1 + t_2) + bit_2 + cit_1 - \frac{a}{2}(t_1 + t_2)^2 - \frac{b}{2}t_2^2 - \frac{c}{2}t_1^2 + \ldots$$

$$= \log\Big\{1 + i(a + b)t_2 + i(a + c)t_1 - \frac{a + c}{2}t_1^2 - \frac{a + b}{2}t_2^2 - at_1t_2 \ldots\Big\} \qquad (5.64)$$

From this it is seen that the mean of the variate corresponding to the occurrence of P's is $n(a + c)$, and that of the variate corresponding to the occurrence of the R's, $n(a + b)$. From the terms in t_1^2 and t_2^2 in the expansion of (5.64) we find also that the variances are $n(a + c)(1 - \overline{a + c})$ and $n(a + b)(1 - \overline{a + b})$. If we now transfer the origin to the mean of the variates we have

$$\log\phi = -\frac{n}{2}\{t_1^2(a + c)(1 - \overline{a + c}) + t_2^2(a + b)(1 - \overline{a + b}) + 2t_1t_2(a - \overline{a + c}\,\overline{a + b})\} + O(n).$$

Thus when the distribution is expressed in standard measure and n allowed to tend to infinity the characteristic function tends to the form

$$\log\phi = -\tfrac{1}{2}(t_1^2 + t_2^2 + 2\rho t_1 t_2) \qquad (5.65)$$

where

$$\rho = \frac{a - (a + c)(a + b)}{\{(a + c)(1 - \overline{a + c})(a + b)(1 - \overline{a + b})\}^{\frac{1}{2}}}$$

This, as was seen in Example 3.15, is the characteristic function of the bivariate form

$$dF = \frac{1}{2\pi(1 - \rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1 - \rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\right\}dx_1\,dx_2 \quad -\infty \leqslant x_1, x_2 \leqslant \infty \qquad (5.66)$$

Thus the multinomial form (5.62) tends to the form (5.66), which may be regarded as the bivariate analogue of the normal distribution.

If the two attributes P and R are independent in the population, that is to say, the proportion of P's among R's is the same as among the not-R's, we have

$$\frac{a}{a+b} = \frac{c}{c+d}$$

and hence

$$\frac{a+c}{a+b+c+d} = \frac{a}{a+b} = \frac{a+c}{1},$$

so that $a - (a+b)(a+c) = 0$. Thus ρ in equation (5.65) vanishes. In this case and only in this case the distribution (5.66) becomes

$$dF = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x_1^2}\, dx_1 \; \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x_2^2}\, dx_2,$$

i.e. x_1 and x_2 are independent variables. This is what we should expect and, indeed, is necessary if our use of the word "independent" in relation to attributes and frequency-distributions is to be consistent.

NOTES AND REFERENCES

For further formulae about the constants of the binomial distribution, including the incomplete moments $\sum_{j=0}^{r-1}\binom{n}{j}q^{n-j}p^j$, see Frisch (1926) and Romanovsky (1925). Some of the results are given as exercises below. See also Haldane (1939). For the formulae of the hypergeometric distribution see K. Pearson (1895 and 1924). On the distribution functions of the binomial and the hypergeometric, see Camp (1924 and 1925).

On Poisson's distribution reference may be made to Whitaker (1914), "Student" (1907 and 1919) and Morant (1921).

Camp, B. H. (1924), "Probability integrals for the point binomial," *Biometrika*, **16**, 163.

—— (1925), "Probability integrals for the hypergeometrical series," *Biometrika*, **17**, 61.

Frisch, R. (1926), see refs. to Chapter 3.

Haldane, J. B. S. (1939), "The cumulants and moments of the binomial," *Biometrika*, **31**, 392.

Morant, G. (1921), "On random occurrences in space and time when followed by a closed interval," *Biometrika*, **13**, 309.

Pearson, K. (1895), "Skew variation in homogeneous material," *Phil. Trans.* A., **186**, 343.

—— (1924*a*), "On the moments of the hypergeometrical series," *Biometrika*, **16**, 157.

—— (1924*b*), "On a certain double hypergeometrical series and its representation by continuous frequency surfaces," *Biometrika*, **16**, 172.

Romanovsky, V. (1925), "On the moments of the hypergeometrical series," *Biometrika*, **17**, 57.

Sheppard, W. F. (1939), *The Probability Integral*, British Association Mathematical Tables, vol. 7, Cambridge University Press.

Soper, H. E. (1922), *Frequency Arrays*, Cambridge University Press.

"Student" (1907), "On the error of counting with a haemacytometer," *Biometrika*, **5**, 351.

—— (1919), "An explanation of deviations from Poisson's law in practice," *Biometrika*, **12**, 211.

Whitaker, L. (1914), "On Poisson's law of small numbers," *Biometrika*, **10**, 36.

EXERCISES

5.1. Show that for the binomial distribution $(q + p)^n$

$$\kappa_{r+1} = pq\frac{d\kappa_r}{dp}, \qquad r > 1.$$

Hence, writing $c = npq$, $g = q - p$, that—

$\kappa_2 = c$; $\kappa_3 = cg$; $\kappa_4 = c - 6c^2$; $\kappa_5 = g\ (c - 12c^2)$; $\kappa_6 = c - 30c^2 + 120c^3$; $\kappa_7 = g\ (c - 60c^2 + 360c^3)$; $\kappa_8 = c - 126c^2 + 1680c^3 - 5040c^4$.

(Cf. Haldane (1939), who gives formulae up to κ_{12}.)

5.2. Show that for the *incomplete* moments about the mean of the binomial

$$\mu_r = \sum_{j=\rho}^{n} (j - np)^r \binom{n}{j} p^j q^{n-j}$$

equation (5.12) holds, i.e.

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right).$$

(Romanovsky, 1925.)

5.3. Writing $T_j = \binom{n}{j} p^j q^{n-j}$, show that the incomplete moments of the binomial are given by

$$\mu_0 = \sum_{j=\rho}^{n} T_j$$

$$\mu_1 = \rho q T_\rho$$

$$\mu_2 = \rho q T_\rho \{\rho - (n+1)p\} + npq\mu_0$$

$$\mu_3 = \rho q T_\rho [\{\rho - (n+1)p\}^2 + pq\ (2n - 1) + npq\ (q - p)\ \mu_0],$$

and generally

$$\mu_r = \rho q T_\rho (\rho - np)^{r-1} + npq \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j - p \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_{j+1}.$$

(Frisch, 1926. This is the generalisation of equation (5.11) to incomplete moments.)

5.4. Show that about the origin of the hypergeometric distribution

$$\mu'_{[r]} = \frac{n^{[r]}(Np)^{[r]}}{N^{[r]}}.$$

5.5. From equation (5.48) derive the recurrence formula for the moments of the binomial

$$\{(1 + E)^r - E^r\}(npq\mu_0 - p\mu_1) = \mu_{r+1}$$

and that for the Poisson distribution

$$\{(1 + E)^r - E^r\}\lambda\mu_0 = \mu_{r+1}.$$

(K. Pearson, 1924.)

5.6. Show that if $y = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}$

$$\int_{-\infty}^{\infty} y^2\,dx = \frac{}{2\sigma\sqrt{\pi}}.$$

Hence, if a normal distribution is grouped in intervals with total frequency N_1, and N_2 is the sum of the squares of frequencies, an estimate of σ is

$$\frac{N_1^2}{2N_2\sqrt{\pi}} = 0{\cdot}282{,}095\frac{N_1^2}{N_2}.$$

For the height data of Table 1.7 show that this gives an estimate of σ equal to 2·553, an error of about 1 per cent.

(Yule (1938), *Biometrika*, **30**, 1.)

5.7. If a distribution of type (5.24) is represented approximately by a binomial $(Q + P)^\nu$, show that

$$\nu P = n\bar{p}$$
$$\nu PQ = n\bar{p}\bar{q} - n \text{ var } p$$

so that $P = \bar{p} + \dfrac{\text{var } p}{\bar{p}}$ and hence is positive ; consequently that ν is positive.

If, however, the distribution is of type (5.28), then

$$P = \bar{p} - \frac{(n - 1) \text{ var } p}{\bar{q}}$$

so that P, and hence ν, may be negative.

(" Student," 1919.)

5.8. The bivariate Poisson series. Show that when a, b and c in equation (5.62) are small but $na(=\lambda_3)$, $nb(=\lambda_1 - \lambda_3)$ and $nc(=\lambda_2 - \lambda_3)$ are finite, the distribution tends to the form whose general term is

$$\frac{\lambda_3^{i}(\lambda_1 - \lambda_3)^{j}(\lambda_2 - \lambda_3)^{k}}{i!\,j!\,k!}\, e^{-\lambda_1 - \lambda_2 + \lambda_3}.$$

5.9. Show that if the frequencies of two symmetrical binomial forms of degree n are superposed so that the rth term of one is added to the $(r + 1)$ term of the other, the resultant frequencies are those of a symmetrical binomial of degree $(n + 1)$. Deduce that if two normal distributions with the same variance and means differing by a small part of the variance are added together, the resultant distribution is nearly normal.

CHAPTER 6

STANDARD DISTRIBUTIONS—(2)

6.1. In this chapter we continue the account, begun in the last, of the standard distributions of statistical theory. From the variety of forms assumed by the frequency-distributions of experience, as exemplified in Chapter 1, it is evident that an elastic system would be required to describe them all in mathematical terms. Three approaches will be considered herein: the first, due to Karl Pearson, seeks to ascertain a *family* of curves which will satisfactorily represent practical distributions; the second, due to Bruns, Gram and Charlier, seeks to represent a given frequency function as a series of derivatives of the normal frequency function; the third, due to Edgeworth, seeks for a transformation of the variate which will throw the distribution at least approximately into the normal form.

Pearson Distributions

6.2. It was noted in **5.23** that in the limiting case the hypergeometric series can be expressed in the form

$$\frac{df}{dx} = \frac{(x-a)f}{b_0 + b_1 x + b_2 x^2} \qquad (6.1)$$

This equation may be considered from a slightly different standpoint. The unimodal distributions of Chapter 1 suggest that it might be worth while examining the class of frequency functions which (*a*) have a single mode, so that $\frac{df}{dx}$ vanishes at some point $x = a$; (*b*) have smooth contact with the x-axis at the extremities, so that $\frac{df}{dx}$ vanishes when $f = 0$. Evidently these conditions are in general obeyed by any distribution of the family (**6.1**). In actual fact, as will be seen below, there are also solutions of (6.1) in particular cases which are J- or U-shaped.

The family of frequency functions defined by (6.1) are known as Pearson distributions. Before obtaining explicit solutions of the equation, we consider certain general results which are true of all members of the system. We have immediately

$$(b_0 + b_1 x + b_2 x^2)\, df = (x - a) f\, dx$$

or

$$x^n (b_0 + b_1 x + b_2 x^2) \frac{df}{dx}\, dx = x^n (x - a) f\, dx.$$

Integrating the left-hand side by parts over the range of the distribution, we find, assuming that the integrals exist,

$$\Big[x^n (b_0 + b_1 x + b_2 x^2) f\Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \{n b_0 x^{n-1} + (n+1) b_1 x^n + (n+2) b_2 x^{n+1}\} f\, dx$$

$$= \int_{-\infty}^{\infty} x^{n+1} f\, dx - a \int_{-\infty}^{\infty} x^n f\, dx \;. \qquad (6.2)$$

Let us assume that the expression in square brackets vanishes at the extremities of the

distribution, i.e. that $\lim_{x \to \pm\infty} x^{n+2}f = 0$ if the range is infinite. We then have, substituting moments for integrals in (6.2):—

$$-nb_0\mu'_{n-1} - (n+1)b_1\mu'_n - (n+2)b_2\mu'_{n+1} = \mu'_{n+1} - a\mu'_n$$

or

$$nb_0\mu'_{n-1} + \{(n+1)b_1 - a)\}\mu'_n + \{(n+2)b_2 + 1\}\mu'_{n+1} = 0 \quad . \quad . \quad (6.3)$$

This equation permits of the determination of any moment from those of lower orders. In fact, all moments can be expressed in terms of a, b_0, b_1 and b_2 and the moments $\mu_0(=1)$ and μ'_1. Conversely we can express these four constants in terms of the moments μ'_1 to μ'_4, or the three moments about the mean μ_2 to μ_4. Putting $n = 0, 1, 2, 3$, successively in (6.3), we find equations for a, b_0, b_1, b_2 which result in

$$\left.\begin{aligned}
a &= -\frac{\mu_3(\mu_4 + 3\mu_2^2)}{A} &&= -\frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2+3)}{A'} \\
b_0 &= -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{A} &&= -\frac{\mu_2(4\beta_2 - 3\beta_1)}{A'} \\
b_1 &= -\frac{\mu_3(\mu_4 + 3\mu_2^2)}{A} &&= -\frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2+3)}{A'} \\
b_2 &= -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{A} &&= -\frac{(2\beta_2 - 3\beta_1 - 6)}{A'}
\end{aligned}\right\} \quad (6.4)$$

where

$$\left.\begin{aligned} A &= 10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2 \\ A' &= 10\beta_2 - 18 - 12\beta_1 \end{aligned}\right\} \quad . \quad (6.5)$$

It follows that a curve of the family (6.1) is completely determined by its first four moments, μ'_1 to μ'_4.

6.3. In equation (6.1) the mode is evidently at the point $x = a$. We have then for the Pearson measure of skewness (3.31)

$$\text{Sk} = \frac{-a}{\sqrt{\mu_2}} = \frac{\sqrt{\beta_1}(\beta_2+3)}{10\beta_2 - 12\beta_1 - 18} \quad . \quad . \quad (6.6)$$

the form given in **3.31**.

Further, if we take an origin at the mode so that $a = 0$ we find

$$\frac{d^2f}{dx^2} = \frac{d}{dx}\frac{xf}{b_0 + b_1x + b_2x^2} = \frac{f}{(b_0 + b_1x + b_2x^2)^2}\{(1-b_2)x^2 + b_0\} \quad (6.7)$$

Thus any points of inflection in the frequency curve are given by

$$x^2 = \frac{b_0}{b_2 - 1} \quad (6.8)$$

Hence there cannot be more than two of them, and if they exist, they are equidistant from the mode. It is not to be inferred that a curve of the family cannot have a single point of inflection, for one point corresponding to the solution of (6.8) may be outside the permissible range of x.

6.4. By a simple transformation of the origin to the mode, (6.1) may be written

$$\left.\begin{aligned} \frac{d}{dx}(\log f) &= \frac{x-a}{B_0 + B_1(x-a) + B_2(x-a)^2} \\ \text{or} \quad \frac{d}{dX}(\log f) &= \frac{X}{B_0 + BX + B_2X^2} \end{aligned}\right\} \quad . \quad . \quad . \quad (6.9)$$

The explicit expression of the frequency function f is thus a matter of integrating the right-hand side of (6.9).

Following Pearson, we may distinguish three main types according as the denominator on the right in (6.9) has real roots of opposite sign, real roots of the same sign, or imaginary roots. Pearson also distinguished ten other types, some entirely trivial, when the B's take particular values.*

Type I

6.5. Let

$$B_0 + B_1X + B_2X^2 = B_2(X + \alpha_1)(X - \alpha_2), \qquad \alpha_1,\ \alpha_2 > 0$$

Then
$$\frac{d}{dX}(\log f) = \frac{X}{B_2(X + \alpha_1)(X - \alpha_2)}$$

$$- \frac{\alpha_1}{B_2(\alpha_1 + \alpha_2)} \cdot \frac{}{(X + \alpha_1)} + \frac{\alpha_2}{B_2(\alpha_1 + \alpha_2)} \cdot \frac{1}{(X - \alpha_2)}$$

giving
$$f = k(X + \alpha_1)^{\frac{\alpha_2}{B_2(\alpha_1+\alpha_2)}}(X - \alpha_1)^{\frac{\alpha_2}{B_2(\alpha_1+\alpha_2)}} \qquad . \quad . \qquad (6.10)$$

This is generally written in the form

$$f = k\left(1 + \frac{x}{a_1}\right)^{m_1}\left(1 - \frac{x}{a_2}\right)^{m_2} \qquad (6.11)$$

where
$$\frac{m_1}{a_1} = \frac{m_2}{a_2}.$$

The range of the curve is from $-a_1$ to a_2 and by integrating between these values we find

$$1 = k\int_{-a_1}^{a_2}\left(1 + \frac{x}{a_1}\right)^{m_1}\left(1 - \frac{x}{a_2}\right)^{m_2} dx,$$

which, on putting $x = (a_1 + a_2)y - a_1$ reduces to

$$1 = k\int_0^1 y^{m_1}(1 - y)^{m_2} . \frac{(a_1 + a_2)^{m_1+m_2+1}}{a_1^{m_1}a_2^{m_2}}$$

$$\frac{k(a_1 + a_2)^{m_1+m_2+1}}{a_1^{m_1}a_2^{m_2}} B(m_1 + 1,\ m_2 + 1).$$

This determines k and we have

$$f = \frac{a_1^{m_1}a_2^{m_2}}{(a_1 + a_2)^{m_1+m_2+1}B(m_1 + 1,\ m_2 + 1)}\left(1 + \frac{x}{a_1}\right)^{m_1}\left(1 - \frac{x}{a_2}\right)^{m_2} . \qquad . \ (6.12)$$

The origin here is the mode. Taking an origin at the start of the curve we have

$$f = \frac{a_1^{m_1}a_2^{m_2}}{(a_1 + a_2)^{m_1+m_2+1}B(m_1 + 1,\ m_2 + 1)} x^{m_1}\left(1 - \frac{x - a_1}{a_2}\right)^{m_2}$$

or again, measuring in units $(a_1 + a_2)$ times the original,

$$f = \frac{1}{B(m_1 + 1,\ m_2 + 1)} x^{m_1}(1 - x)^{m_2} \qquad . \ (6.13)$$

6.6. In these expressions the a's are necessarily positive, but the m's may have any value not less than -1. They cannot be less because the distribution function of (6.12) or (6.13) would not then converge.

* The numbering of the types followed herein is that of Elderton (1938). Some variations occur in earlier literature and the reader must not be surprised to find the normal curve referred to occasionally as Type VII.

If m_1, $m_2 > 0$ the distribution is evidently unimodal and zero at its extremities. If one of the m's is between 0 and 1 the corresponding terminal frequency is still zero, but the frequency curve makes a sharp angle with the x-axis, for $\frac{df}{dx}$ is not zero at the terminal. If one and only one m is less than zero the curve has an infinite ordinate and is thus J-shaped. If both m's are less than zero the curve is U-shaped.

The condition that $B_0 + B_1X + B_2X^2$ shall have real roots of opposite sign is that B_0 and B_2 are of opposite sign, which is equivalent to

$$\frac{B_1^2}{4B_0B_2} < 0$$

or, in terms of β_1 and β_2, from (6.4),

$$\frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} < 0. \qquad (6.14)$$

The quantity on the left was denoted by Pearson by the letter κ and provides a criterion which will occur again below.

The frequency function of the Type I curve is calculable directly from its equation. The distribution function, as may be seen from (6.13), is expressible in terms of incomplete B-functions.

Type VI

6.7. If the roots of $B_0 + B_1X + B_2X^2$ are real and of the same sign it is easy to see, in the manner of the preceding sections, that the frequency functions may be written in the form

$$f = \frac{a^{q_1-q_2-1}}{B(q_1 - q_2 - 1,\ q_2 + 1)} x^{-q_1}(x - a)^{q_2} \qquad \text{where } q_1 > q_2 - 1 \qquad (6.15)$$

where the range lies from a to ∞ if a is positive and from $-\infty$ to a if a is negative. By the simple transformation $y = \frac{a}{x}$ this reduces to the Type I form (6.13).

It will readily be verified that if $q_2 > 0$ the curves are unimodal with zero frequencies at the terminals. If $q_2 < 0$ the start is J-shaped and the distribution falls away to zero at infinity. The distribution function may be expressed in terms of incomplete B-functions, and in this case the quantity κ of (6.14) is greater than unity.

Type IV

6.8. If the roots of $B_0 + B_1X + B_2X^2$ are imaginary we have

$$\frac{d}{dX}(\log f) = \frac{X}{B_2\left\{\left(X + \frac{B_1}{2B_2}\right)^2 + \frac{B_0}{B_2} - \frac{B_1^2}{4B_2^2}\right\}}$$

$$= \frac{X}{B_2\{(X+\gamma)^2 + \delta^2\}}, \text{ say}$$

giving

$$\log f = \log k + \frac{1}{2B_2}\log\{(X+\gamma)^2 + \delta^2\} - \frac{\gamma}{B_2\delta}\tan^{-1}\frac{X+\gamma}{\delta}$$

$$f = k\{(X+\gamma)^2 + \delta^2\}^{\frac{1}{2B_2}} \exp\left\{-\frac{\gamma}{B_2\delta}\tan^{-1}\frac{X+\gamma}{\delta}\right\}. \qquad (6.16)$$

This is Pearson's Type IV and is usually written in the form

$$f = k\left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-\nu \tan^{-1} \frac{x}{a}} \quad . \quad . \quad . \quad . \quad . \quad (6.17)$$

The distribution has unlimited range in both directions, tends to zero at infinity and is unimodal. The calculation of its ordinates may be assisted by some tables by Comrie (1939). The distribution function has to be found either by quadratures from the frequency function or by the use of some tabulated integrals given in *Tables for Statisticians and Biometricians*, Part I. For instance, for the constant k in (6.17) we have

$$\frac{1}{k} = \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-\nu \tan^{-1} \frac{x}{a}} dx$$

$$= a\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{2m-2} \theta \, e^{-\nu\theta} \, d\theta$$

$$= aF(2m - 2, \nu) \text{ in Pearson's notation.}$$

In this case the quantity κ of (6.14) lies between 0 and 1.

The above are the three main types in the Pearsonian system. The remaining types are described briefly below. A number of results which the reader can easily verify for himself are given without proof.

The Normal Distribution

6.9. If, in equation (6.9), $a = B_1 = B_2 = 0$, we have

$$\frac{d}{dx}(\log f) = \frac{x}{B_0}$$

$$\log f = \int \frac{x}{B_0} dx = \frac{x^2}{2B_0} + \log k$$

$$f = ke^{\frac{x^2}{2B_0}}.$$

If this frequency function is to have a convergent distribution function, B_0 must be negative, $= -\sigma^2$ say, and we get the familiar form

$$f = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{x^2}{2\sigma^2}} \qquad -\infty \leqslant x \leqslant \infty.$$

Thus the normal distribution itself is one of Pearson's types.

Type II

6.10. If in equation (6·9) $B_1 = 0$ and B_2, B_0 are of opposite signs, the distribution, a particular case of Type I, becomes of the character

$$f = \frac{1}{aB(\frac{1}{2}, m + 1)}\left(1 - \frac{x^2}{a^2}\right)^m \qquad -a \leqslant x \leqslant a \qquad (6.18)$$

(a here being different from the a of (6.9)).

In this case the criterion κ of equation (6.14) is zero. The distribution is symmetrical about the origin and ranges from $-a$ to $+a$. If $m > 0$ it is unimodal with contact at

the terminals of the range; if $m < 0$ it is U-shaped. If $m = 0$ the distribution becomes

$$f = \frac{1}{2a}, \qquad -a \leqslant x \leqslant a \quad . \tag{6.19}$$

the so-called "rectangular" distribution.

Type VII

6.11. If in (6.9) $B_1 = 0$ and B_2, B_0 are of the same sign, we find

$$f = \frac{1}{aB(\frac{1}{2}, m - \frac{1}{2})}\left(1 + \frac{x^2}{a^2}\right)^{-m} \qquad -\infty \leqslant x \leqslant \infty \quad . \quad . \tag{6.20}$$

The range is now unlimited in both directions. Here also the criterion κ of (6.14) vanishes, but the difference between this case and that of Type II lies in the fact that here $\beta_2 > 3$, whereas in the Type II case $\beta_2 < 3$.

Type III

6.12. If in (6.9) $B_2 = 0$ we obtain the distribution

$$f = \frac{p^{p+1}}{ae^p\Gamma(p+1)}\left(1 + \frac{x}{a}\right)^p e^{-\frac{\gamma}{a}} \quad . \tag{6.21}$$

this being the form with the origin at the mode. The curve is unlimited in one direction (positive or negative as $\frac{p}{a}$ is positive or negative). It is unimodal if $p > 0$, J-shaped if $p < 0$. The condition $B_2 = 0$, from (6.4), is equivalent to $2\beta_2 - 3\beta_1 - 6 = 0$, i.e. κ of (6.14) is infinite.

Type V

6.13. If the roots of $B_0 + B_1X + B_2X^2$ are equal, i.e. $\kappa = 1$, we arrive at the distribution

$$f = \frac{\gamma^{p-1}}{\Gamma(p-1)}x^{-p}e^{-\frac{\gamma}{x}} \qquad 0 \leqslant x \leqslant \infty \tag{6.22}$$

which ranges from 0 to ∞ and is unimodal.

Types VIII, IX, X, XI and XII

6.14. The remaining types are of a more special character still.

If in (6.9) $B_0 = 0$, $B_1 > 0$ we have

$$\textit{Type VIII}: \ f = \frac{1-m}{a} \ 1 + \frac{x}{a} \qquad 0 \leqslant m \leqslant 1, \quad -a \leqslant x \leqslant 0 \quad . \tag{6.23}$$

If $B_0 = 0$, $B_1 < 0$ we have

$$\textit{Type IX}: \ f = \frac{m+1}{a} \ 1 + \frac{x}{a} \qquad -a \leqslant x \leqslant 0 \tag{6.24}$$

If $B_0 = B_2 = 0$ we have

$$\textit{Type X}: \ f = \frac{1}{\sigma} e^{-\frac{x}{\sigma}} \qquad 0 \leqslant x \leqslant \infty \tag{6.25}$$

If $B_0 = B_1 = 0$ we have

$$\textit{Type XI}: \ f = b^{m-1}(m-1)x^{-m} \qquad b \leqslant x \leqslant \infty \tag{6.26}$$

Finally, as a particular case of Type I when $5\beta_2 - 6\beta_1 - 9 = 0$, equations (6.4) become indeterminate. In this case we have

$$Type\ XII:\ f = \left(\frac{a_1}{a_2}\right)^m \cdot \frac{1}{(a_1 + a_2)B(1+m, 1-m)}\left(\frac{1 + \dfrac{x}{a_1}}{1 - \dfrac{x}{a_2}}\right)^m \qquad |m| < 1,$$
$$-a_1 \leqslant x \leqslant a_2 \qquad . \quad (6.27)$$

6.15. Pearson curves of Types I and III, and to a somewhat smaller extent, those of Types V and VII, arise in the theory of sampling and would in any case have to be studied in that theory. Apart from this, the principal use that has been made of the distributions in the theory of statistics is in fitting them to observed distributions such as those of Chapter 1. It has been found that in many cases the Pearson distributions provide a remarkably good fit to observation.

A systematic account of the technique of fitting will be found in Elderton's *Frequency Curves and Correlation* (1938). We will here merely indicate the general principles and give one example of fitting in what is, perhaps, the most difficult case.

6.16. All the Pearson distributions are determined by the first four moments, μ_1' to μ_4' inclusive, except some of the degenerate types which are determined by fewer than four moments. Pearson's method of fitting consists of

(1) determining the numerical values of the first four moments of the observed distribution;
(2) calculating the numerical values of β_1, β_2, κ (equation 6.14) and hence determining the type to which the distribution belongs;
(3) equating the observed moments to the moments of the appropriate distribution expressed in terms of its parameters; and
(4) solving the resulting equations for those parameters, whereupon the distribution is determined.

The following example will illustrate the process :—

Example 6.1

In Table 1.15 there are shown, in the column totals, a distribution of 9440 beans according to length. The figures are repeated in Table 6.1 on page 150. Required to fit a Pearson distribution to these data.

For the moments it is found that, with Sheppard's corrections,

$$\begin{aligned} \mu_1 \text{ (centre at } 14{\cdot}5) &= -\,0{\cdot}190{,}783{,}898 \\ \mu_2 &= \quad 3{\cdot}238{,}424{,}951 \\ \mu_3 &= -\,5{\cdot}306{,}566{,}352 \\ \mu_4 &= \quad 50{\cdot}999{,}624{,}044 \end{aligned}$$

$$\beta_1 = 0{\cdot}829{,}135{,}838, \qquad \sqrt{\beta_1} = -\,0{\cdot}910{,}569$$
$$\beta_2 = 4{\cdot}862{,}944{,}362$$

First of all, as to type. For the criterion κ (6.14) we have

$$\kappa = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}$$
$$= \frac{51{\cdot}262}{84{\cdot}040}.$$

This lies between 0 and 1 and hence the appropriate curve is Type IV. We have to determine a, m, ν in

$$f = \frac{N}{aF(2m-2,\nu)}\left(1+\frac{x^2}{a^2}\right)^{-m} e^{-\nu\tan^{-1}\frac{x}{a}}.$$

Writing $\tan\phi = \frac{x}{a}$ and $2m-2 = r$ we find

$$\mu'_n = k\int_{-\infty}^{\infty} a^{n+1}\cos^{r-n}\theta\sin^n\theta\, e^{-\nu\theta}\,d\theta,$$

whence, integrating by parts with $\cos^{r-n}\theta\sin\theta$ as one part,

$$\mu'_n = \frac{a}{r-n+1}\{(n-1)a\mu'_{n-2} - \nu\mu'_{n-1}\},$$

a particular case of (6.3). Hence, in terms of moments about the mean,

$$\mu'_1 = -\frac{a\nu}{r}$$

$$\mu_2 = \frac{a^2}{r^2(r-1)}(r^2+\nu^2)$$

$$\mu_3 = -\frac{4a^3\nu(r^2+\nu^2)}{r^3(r-1)(r-2)}$$

$$\mu_4 = \frac{3a^4(r^2+\nu^2)\{\overline{r+6}(r^2+\nu^2)-8\nu^2\}}{r^4(r-1)(r-2)(r-3)}$$

whence it is found that

$$r = \frac{6(\beta_2-\beta_1-1)}{2\beta_2-3\beta_1-6}$$

$$\nu = -\frac{r(r-2)\sqrt{\beta_1}}{\sqrt{\{16(r-1)-\beta_1(r-2)^2\}}}$$

$$a = \sqrt{\left[\frac{\mu_2}{16}\{16(r-1)-\beta_1(r-2)^2\}\right]}.$$

Substituting for β_2, β_1 and μ_2 we find

$$r = 14{\cdot}697{,}72, \qquad m = 8{\cdot}348{,}86$$
$$\nu = 18{\cdot}380{,}43 \qquad a = 4{\cdot}159{,}49$$

The signs here want a little watching. r and m present no difficulty; but a is to be taken positive and ν positive since $\sqrt{\beta_1}$ is to be considered negative.

From the tables of $F(r,\nu)$ we evaluate the constant term k and finally arrive at

$$f = 0{\cdot}395{,}121\left(1+\frac{x^2}{17{\cdot}301{,}34}\right)^{-8{\cdot}348{,}86} e^{-18{\cdot}380{,}43\tan^{-1}\frac{x}{4{\cdot}159{,}49}}.$$

The frequencies given by this curve are shown in Table 6.1 on page 150.

6.17. The following points are worth noting in connection with the fitting of Pearson curves to observational data:—

(1) Although the various types have dissimilar analytical equations they merge into one another in geometrical shape. For instance, Type V may be regarded as transitional

between Types IV and VI and is very similar to the shape assumed by those curves near $\kappa = 1$.

(2) It is tacitly assumed that the data can be represented by a curve with finite moments up to the fourth order at least. Curves for which higher moments do not exist were called by Pearson heterotypic; but there is nothing sinister about them except that they do not fall within the Pearsonian system.

(3) In calculating moments, Sheppard's corrections are usually to be employed when there are contacts of sufficiently high order at the terminals. In the case of J- or U-distributions the other corrections mentioned in **3.27** may be employed. This case sometimes raises difficulties in that the resultant curve does not start in the right place. In such circumstances there is no golden rule. The most satisfactory course is to try several curves (or the same curve translated to several points) and to judge by the results which of them gives the best fit.

(4) The quadrature of Pearson curves, as indicated in the foregoing, may in some cases be effected by tabulated integrals; but the more generally applicable procedure appears to be to calculate ordinates direct from the equation of the curve and then to find areas in ranges by Simpson's rule, Weddle's rule, or some similar process of quadrature.

6.18. The mathematical description of an observed distribution by a Pearson curve may be regarded from two rather different standpoints. If our object (for instance in actuarial work) is to obtain a mathematical expression which will satisfactorily represent observation and allow of accurate graduation and interpolation, fitting by moments is generally satisfactory. The method has, however, been criticised when the observed data are regarded as samples from a population, and it is desired to find a mathematical representation *of that population*. In such cases the moments calculated from observation are only estimates of population-moments. It has been objected that they may be inefficient estimates, and alternative methods have been proposed. We shall have to defer a full discussion of this point until the second volume.

6.19. Other systems of curves have been studied, mainly by Scandinavian writers, with a view to representing frequency functions by expansions in series. It is well known in mathematical and physical work that functions can often be usefully expressed as a series of terms such as powers of the variable (Taylor's series) or trigonometrical functions (Fourier's series). Neither of these forms is very suitable for frequency functions, but we proceed to consider another set of functions with more promising possibilities.

Tchebycheff-Hermite Polynomials

6.20. Writing

$$\alpha(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{x^2}{2}}$$

and

$$D = \frac{d}{dx}$$

consider successive derivatives of $\alpha(x)$ with respect to x. We have

$$\begin{aligned} D\alpha(x) &= -x\alpha(x) \\ D^2\alpha(x) &= (x^2 - 1)\alpha(x) \\ D^3\alpha(x) &= (3x - x^3)\alpha(x). \end{aligned}$$

and so on. The result will obviously be, in general, a polynomial in x multiplied by $\alpha(x)$. We then define the Tchebycheff-Hermite polynomial $H_r(x)$ by the equation

$$(-D)^r\alpha(x) = H_r(x)\alpha(x) \tag{6.28}$$

Evidently $H_r(x)$ is of degree r in x and the coefficient of x^r is unity. By convention $H_0 = 1$. We have

$$\alpha(x-t) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{x^2}{2} + tx - \frac{t^2}{2}\right) = \alpha(x) \exp\left(tx - \frac{t^2}{2}\right)$$

and also, by Taylor's theorem

$$\alpha(x-t) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} t^j D^j \alpha(x) = \sum_{j=0}^{\infty} \frac{t^j}{j!} H_r(x)\alpha(x).$$

Consequently $H_r(x)$ is the coefficient of $\frac{t_r}{r!}$ in $\exp\left(tx - \frac{t^2}{2}\right)$. It follows that

$$H_r(x) = x^r - \frac{r^{[2]}}{2\ 1!}x^{r-2} + \frac{r^{[4]}}{2^2 2!}x^{r-4} - \frac{r^{[6]}}{2^3 3!}x^{r-6} \tag{6.29}$$

The first ten polynomials are

$$\left.\begin{aligned}
H_0 &= 1\\
H_1 &= x\\
H_2 &= x^2 - 1\\
H_3 &= x^3 - 3x\\
H_4 &= x^4 - 6x^2 + 3\\
H_5 &= x^5 - 10x^3 + 15x\\
H_6 &= x^6 - 15x^4 + 45x^2 - 15\\
H_7 &= x^7 - 21x^5 + 105x^3 - 105x\\
H_8 &= x^8 - 28x^6 + 210x^4 - 420x^3 + 105\\
H_9 &= x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x\\
H_{10} &= x^{10} - 45x^8 + 630x^6 - 3150x^4 + 4725x^2 - 945
\end{aligned}\right\} \tag{6.30}$$

6.21. The polynomials have a number of interesting properties. Differentiating the identity

$$\exp\left(-\frac{t^2}{2}\right) = \sum_{j=0}^{\infty} \frac{t^j H_j(x)}{j!}$$

with respect to x and identifying coefficients in t^r we have

$$\frac{d}{dx}H_r(x) = rH_{r-1}(x) \tag{6.31}$$

and generally

$$D^j H_r(x) = r^{[j]} H_{r-j}(x) \tag{6.32}$$

Differentiating the identity with respect to t and identifying coefficients in t^{r-1} we have

$$H_r(x) - xH_{r-1}(x) + (r-1)H_{r-2}(x) = 0. \tag{6.33}$$

From (6.31) and (6.33) together we find

$$\frac{d^2H_r(x)}{dx^2} - x\frac{dH_r(x)}{dx} + rH_r(x) = 0. \tag{6.34}$$

It is also known that the equation in x, $H_r(x) = 0$, has r real roots, each not greater in absolute value than $\sqrt{\frac{r(r-1)}{2}}$. (Cf. Charlier, 1931.)

Tables of the values of the first six polynomials to 10 decimal places proceeding by $x = 0\ (0{\cdot}01)\ 4$ have been given by Jørgensen (1916).

6.22. The polynomials have an important orthogonal property, namely, that

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)\alpha(x)\,dx = 0 \qquad m \neq n$$
$$= n! \qquad m = n \qquad . \quad (6.35)$$

In fact, integrating by parts, we have, if $m < n$,

$$\int_{-\infty}^{\infty} H_m H_n \alpha\,dx = (-1)^n \int_{-\infty}^{\infty} H_m D^n \alpha\,dx$$
$$= (-1)^n \Big[H_m D^{n-1}\alpha\Big]_{-\infty}^{\infty} + (-1)^{n-1}\int_{-\infty}^{\infty} \frac{dH_m}{dx} D^{n-1}\alpha\,dx.$$

The term in square brackets vanishes and, in virtue of (6.31), the integral becomes

$$m(-1)^{n-1}\int_{-\infty}^{\infty} H_{m-1}D^{n-1}\alpha\,dx.$$

Continuing the process, we find either zero, if m is not equal to n, or $m!$ if $m = n$.

The Gram-Charlier Series of Type A

6.23. Suppose now that a frequency function can be expanded formally in series of derivatives of $\alpha(x)$. (We shall discuss the conditions under which such an expansion is valid below.) We have then

$$f(x) = \sum_{j=0}^{\infty} c_j H_j(x)\alpha(x).$$

Multiplying by $H_r(x)$ and integrating from $-\infty$ to ∞ we have, in virtue of the orthogonal relationship (6.35),

$$c_r = \frac{1}{r!}\int_{-\infty}^{\infty} f(x)H_r(x)\,dx \qquad . \quad (6.36)$$

The reader familiar with harmonic analysis will recognise the resemblance between this procedure and the evaluation of constants in a Fourier series.

Substituting in (6.36) the explicit value of $H_r(x)$ given in (6.29) we find

$$c_r = \frac{1}{r!}\left\{\mu'_r - \frac{r^{[2]}}{2\,1!}\mu'_{r-2} + \frac{r^{[4]}}{2^2 2!}\mu'_{r-4} - \ldots\right\} \qquad . \quad . \quad . \quad (6.37)$$

In particular, for moments about the mean,

$$\left.\begin{aligned}
c_0 &= 1\\
c_1 &= 0\\
c_2 &= \tfrac{1}{2}(\mu_2 - 1)\\
c_3 &= \tfrac{1}{6}\mu_3\\
c_4 &= \tfrac{1}{24}(\mu_4 - 6\mu_2 + 3)\\
c_5 &= \tfrac{1}{120}(\mu_5 - 10\mu_3)\\
c_6 &= \tfrac{1}{720}(\mu_6 - 15\mu_4 + 45\mu_2 - 15)\\
c_7 &= \tfrac{1}{5040}(\mu_7 - 21\mu_5 + 105\mu_3)\\
c_8 &= \tfrac{1}{40320}(\mu_8 - 28\mu_6 + 210\mu_4 - 420\mu_2 + 105)
\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (6.38)$$

Thus we find the formal expansion

$$f(x) = \alpha(x)\{1 + \tfrac{1}{2}(\mu_2 - 1)H_2 + \tfrac{1}{6}\mu_3 H_3 + \tfrac{1}{24}(\mu_4 - 6\mu_2 + 3)H_4 + \ldots\} \quad . \quad (6.39)$$

If $f(x)$ is in standard measure the series becomes

$$f(x) = \alpha(x)\{1 + \tfrac{1}{6}\mu_3 H_3 + \tfrac{1}{24}(\mu_4 - 3)H_4 + \ldots\} \quad . \quad . \quad . \quad (6.40)$$

This is the so-called Gram-Charlier series of Type A.

Edgeworth's Form of the Type A Series

6.24. Consider the characteristic function of a term $H_r(x)\alpha(x)$.

Since $$\alpha(t) = e^{-\frac{t^2}{2}} = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{x^2}{2}}\,dx$$

we have $$\sqrt{(2\pi)}\frac{d^r}{dt^r}\alpha(t) = (-1)^r\sqrt{(2\pi)}H_r(t)\alpha(t) = \int_{-\infty}^{\infty} i^r x^r \frac{e^{itx}}{\sqrt{(2\pi)}}\,e^{-\frac{x^2}{2}}\,dx$$

and thus the characteristic function of $x^r\alpha(x)$ is $i^r\sqrt{(2\pi)}H(t)\alpha(t)$. Conversely, by the Inversion Theorem of 4.3, we have

$$x^r\alpha(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\, i^r\sqrt{(2\pi)}H(t)\alpha(t)\,dt.$$

Interchanging x and t, we find

$$\sqrt{(2\pi)}(-i)^r t^r\alpha(t) = \int_{-\infty}^{\infty} e^{-ixt}H(x)\alpha(x)\,dx$$

and hence, changing the sign of t, that the characteristic function of $H(x)\alpha(x)$ is $\sqrt{(2\pi)}i^r t^r\alpha(t)$.

Consider now the expression

$$\exp\,(\kappa_r D^r)\alpha(x) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.41)$$

Its characteristic function is

$$\begin{aligned}\int_{-\infty}^{\infty} e^{itx}\exp\,(\kappa_r D^r)\alpha(x)dx &= \int_{-\infty}^{\infty} e^{itx}\Sigma\left(\frac{\kappa_r{}^j D^{rj}}{j!}\right)\alpha(x)\,dx \\ &= \Sigma\frac{\kappa_r{}^j}{j!}\int_{-\infty}^{\infty} e^{itx}D^{rj}\alpha(x)\,dx \\ &= \Sigma\frac{\kappa_r{}^j}{j!}\int_{-\infty}^{\infty} e^{itx}(-1)^{rj}H_{rj}(x)\alpha(x)\,dx \\ &= \Sigma\frac{\kappa_r{}^j}{j!}\sqrt{(2\pi)}(-i)^{rj}t^{rj}\alpha(t) \\ &= \sqrt{(2\pi)}\exp - (\kappa_r it)^r\alpha(t) \qquad (6.42)\end{aligned}$$

In a similar way it will be seen that the characteristic function of

$$\exp\left\{-\frac{\kappa_1 - a}{1!}D + \frac{\kappa_2 - b}{2!}D^2 - \frac{\kappa_3}{3!}D^3 + \frac{\kappa_4}{4!}D^4 \ldots\right\}\alpha(x) \qquad (6.43)$$

is equal to

$$\sqrt{(2\pi)}\alpha(t)\exp\ \frac{\kappa_1 - a}{1!}it + \frac{\kappa_2 - b}{2!}(it)^2 + \frac{\kappa_3}{3!}(it)^3 + \frac{\kappa_4}{4!}(it)^4 + \ldots \qquad (6.44)$$

More generally, if

$$\beta(x) = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}$$

the characteristic function of

$$\exp\left\{-\frac{\kappa_1 - a}{1!}D + \frac{\kappa_2 - b}{2!}D^2 - \frac{\kappa_3}{3!}D^3 + \frac{\kappa_4}{4!}D^4 \ldots\right\}\beta(x) \tag{6.45}$$

is equal to

$$\sqrt{(2\pi)}\alpha(t\sigma)e^{imt}\exp\left\{\frac{\kappa_1 - a}{1!}it + \frac{\kappa_2 - b}{2!}(it)^2 + \frac{\kappa_3}{3!}(it)^3 + \frac{\kappa_4}{4!}(it)^4\right. \tag{6.46}$$

as may be seen by the same line of argument.

Now suppose that (6.45) represents a frequency function. Its cumulative function is then the logarithm of (6.46), i.e. is equal to

$$\frac{(\kappa_1 - a + m)it}{1!} + \frac{\kappa_2 - b + \sigma^2}{2!}(it)^2 + \frac{\kappa_3}{3!}(it)^3 + \frac{\kappa_4}{4!}(it)^4 +$$

and hence its cumulants are $\kappa_1 - a + m$, $\kappa_2 - b + \sigma^2$, κ_3, κ_4 . . ., κ_r . . ., etc. We may take $a = m$ and $b = \sigma^2$ and thus we obtain a distribution whose cumulants are κ_1, κ_2, . . . etc. Now if these are in fact the cumulants of a distribution the series (6.45) must be equal to that distribution, provided that (1) the series converges to a frequency function, and (2) it is uniquely determined by its moments.

If we take the frequency function to be expressed in standard measure, then $\kappa_1 = 0$, $\kappa_2 = 1$ and (6.45) becomes

$$\exp\left\{-\kappa_3\frac{D^3}{3!} + \kappa_4\frac{D^4}{4!} \ldots\right\}\alpha(x) = f(x) \tag{6.47}$$

where we have written $\alpha(x)$ for $\beta(x)$ because now m vanishes and $\sigma^2 = 1$.

A series of this kind was derived by Edgeworth (1904), though from an entirely different approach through the theory of elementary errors. Equation (6.47) is formally identical with (6.40), and the reader who consults the original memoirs on this subject may be puzzled by the fact that Edgeworth claimed his series to be different from the Type *A* series and better as a representation of frequency functions. The explanation is that for practical purposes it is necessary to take only a finite number of terms in the series and to neglect the remainder. If we take the first k terms in (6.40) the result is in general different from that obtained by taking the first $(k - 1)$ terms of the operator in the exponential of (6.47). The argument centred on the fact (cf. Example 6.3 below) that the terms in (6.40) do not tend regularly to zero from the point of view of elementary errors, so that in general no term is negligible compared with a preceding term.

6.25. In standard measure the relations (6.38) become, in terms of cumulants,

$$\begin{aligned}
c_0 &= 1, \quad c_1 = c_2 = 0 \\
c_3 &= \frac{\kappa_3}{6} \\
c_4 &= \frac{\kappa_4}{24} \\
&\ \frac{\kappa_5}{120} \\
c_6 &= \frac{1}{720}(\kappa_6 + 10\kappa_3^2) \\
c_7 &= \frac{1}{5040}(\kappa_7 + 35\kappa_4\kappa_3) \\
c_8 &= \frac{1}{40320}(\kappa_8 + 56\kappa_5\kappa_3 + 35\kappa_4^2)
\end{aligned} \tag{6.48}$$

6.26. In the practical representation of frequency functions by the Type A series only the first few terms can be taken into account. The term in $H_r(x)$ has a coefficient dependent on μ_r and for $r > 4$ this is unreliable owing to sampling fluctuations. When sampling effects are not in question the series may be taken to more terms, usually not higher than the term in H_6. We should then have to investigate how far the observed distribution can be represented by the series

$$\alpha(x)\left(1 + \frac{\kappa_3}{6}H_3 + \frac{\kappa_4}{24}H_4 + \frac{\kappa_5}{120}H_5 + \frac{\kappa_6 + 10\kappa_3^2}{720}H_6\right) \quad . \quad . \quad . \quad (6.49)$$

in the hope that the remainder after these terms could be neglected in comparison.

It may be noted in passing that the distribution function of such a series is easy to obtain. If

$$f(x) = \Sigma a_r H_r(x)\alpha(x),$$

then

$$\int_{-\infty}^{x} f(x) = \Sigma a_r \int_{-\infty}^{x} H_r(x)\alpha(x)\, dx$$
$$= -\Sigma a_r H_{r-1}(x)\,\alpha(x). \quad (6.50)$$

TABLE 6.1

Fitting of Pearson Type IV Distribution and Gram-Charlier Type A Series to the Data of Length of Beans (Table 1.15).

(From Pretorius, 1930.)

Length of Beans (mm.)	Observed Frequency.	Type IV.	Type A. (1)	Type A. (2)	Type A. (3)
—	—				
—	—	—	16·3	− 15·2	2·0
17·0	6	1·4	12·8	13·7	− 35·3
16·5	55	28·5	25·6	116·6	22·3
16·0	275	299·3	241·7	370·4	438·1
15·5	1129	1181·6	1012·7	926·2	1214·0
15·0	2082	2132·6	2155·4	1833·0	1866·9
14·5	2294	2229·8	2593·0	2506·4	2112·8
14·0	1787	1638·9	1788·4	2082·6	1916·7
13·5	929	968·9	713·4	921·3	1183·4
13·0	437	503·6	280·7	199·0	371·2
12·5	199	243·7	258·7	132·1	66·9
12·0	115	113·8	206·2	178·1	101·2
11·5	70	52·5	98·7	117·0	107·1
11·0	36	24·2	29·6	43·5	54·0
10·5	18	11·3	5·9	10·0	15·4
10·0	7	5·4			
9·5	1	2·6	·9	1·7	3·3
—	—	1·9			
—	—		—	—	—
TOTALS	9440	9440	9440	9440	9440

(The brackets mean that the frequencies shown are rounded up and include some small frequency in blank rows covered by the brackets.)

Example 6.2

Consider the fitting of a Type A series to the bean data of Example 6.1.

We have already found the first four moments. In standard measure we have

$$\mu_3 = -\ 0\cdot910,569$$
$$\mu_4 = 4\cdot862,944$$

and we also find

$$\mu_5 = -\ 12\cdot574,125$$
$$\mu_6 = 53\cdot221,083$$

Hence the series is

$$9440\alpha(x)\{1 - 0\cdot151,762\,H_3 + 0\cdot077,622,7\,H_4 - 0\cdot028,903,6\,H_5 + 0\cdot014,273,5\,H_6\}.$$

Table 6.1, on page 150, due to Pretorius (1930), shows the frequencies given by taking the first three, the first four and the first five terms of this series (columns headed Type A(1), Type A(2) and Type A(3) respectively). A glance at the figures will show that the four- and five-term series is no better than the three-term and, if anything, rather worse. Furthermore, the five-term series gives negative frequencies at one terminal and a mode at 12 mm., which is contrary to the data. The representation is clearly not very satisfactory and no better than that given by the Pearson Type IV curve.

Tetrachoric Functions

6.27. The terms $H_r(x)\alpha(x)$ may be obtained from Jørgensen's tables combined with those of the exponential $e^{-\frac{x^2}{2}}$. Some related functions have also been tabulated in *Tables for Statisticians and Biometricians*, Parts I and II. The function

$$\tau_r = \frac{(-1)^{r-1}D^{r-1}\alpha(x)}{(r!)^{\frac{1}{2}}} = \frac{H_{r-1}(x)\alpha(x)}{(r!)^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad (6.51)$$

is known as the Tetrachoric Function of order r, and tables are available to seven places of decimals for $r = 0$ (1) 30 and $x = 0$ (0·1) 4. In the notation of these functions, series (6.49) would become

$$f(x) = \tau_1(x) + \frac{\sqrt{24}}{6}\kappa_3\tau_4(x) + \frac{\sqrt{120}}{24}\kappa_4\tau_4(x) + \ldots$$

and the particular series of Example 6.2 would be

$$f(x) = 9440\{\tau_1(x) - 0\cdot743,477\,\tau_4(x) + 0\cdot850,313\,\tau_5(x) - 0\cdot775,565\,\tau_6(x) + 1\cdot013,318\,\tau_7(x)\}.$$

The reason for the definition and the name of the function will appear in Chapter 14.

6.28. Up to this point it has been assumed that a frequency function possesses a convergent Type A series. We shall not here enter into a discussion of the conditions under which this is so, except to warn the reader that a great many mistakes have been made on the subject and to quote some theorems without proof.

(1) Cramér (1925). If $f(x)$ is a function which has a continuous derivative such that

$$\int_{-\infty}^{\infty}\left(\frac{df}{dx}\right)^2 e^{\frac{x^2}{2}}\,dx$$

converges and if $f(x)$ tends to zero as $|x|$ tends to infinity, then $f(x)$ may be developed in the series

$$f(x) = \sum_{j=0}^{\infty}\frac{c_j}{j!}D^j\alpha(x), \qquad (6.52)$$

where c_j is given by

$$= \int_{-\infty}^{\infty} f(x) H_j(x)\, dx.$$

This series is absolutely and uniformly convergent for $-\infty \leqslant x \leqslant \infty$.

(2) A theorem by Cramér (1925) based on one by Galbrun. If $f(x)$ is of bounded variation in every finite interval and if

$$\int_{-}^{\infty} |f(x)|\, e^{\frac{-}{4}}\, dx$$

exists, then the expansion of $f(x)$ in the series (6.52) converges everywhere to the sum $\frac{1}{2}\{f(x+0)+f(x-0)\}$. The convergence is uniform in every finite interval of continuity.

Cramér has also shown that this last theorem cannot be substantially improved upon as regards the behaviour of $f(x)$ at infinity. Consider in fact the function $f(x) = e^{-\lambda x^2}$. We have, in virtue of (6.33) and (6.31),

$$\int_{-\infty}^{\infty} e^{-\lambda x^2} H_r(x)\, dx = \int_{-\infty}^{\infty} e^{-\lambda x^2}.x H_{r-1}\, dx - (r-1)\int_{-\infty}^{\infty} e^{-\lambda x^2}.\ H_{r-2}\, dx$$

$$= \left[\frac{e^{-\lambda x^2}}{-2\lambda} H_{r-1}\right]_{-\infty}^{\infty} + \frac{r-1}{2\lambda}\int_{-\infty}^{\infty} e^{-\lambda x^2} H_{r-2}\, dx - (r-1)\int_{-\infty}^{\infty} e^{-\lambda x^2}\ H_{r-2}\, dx$$

$$= (r-1)\left(\frac{1}{2\lambda} - 1\right)\int_{-\infty}^{\infty} e^{-\lambda x^2} H_{r-2}(x)\, dx.$$

If r is odd the integral vanishes because H_r is an odd function of x. If r is even, say $2r$, the integral becomes

$$(2r-1)(2r-3)\ .\ .\ .\ 1\left(\frac{1}{2\lambda} - 1\right)^{\nu}.\frac{1}{\sqrt{(2\lambda)}}$$

$$= \frac{1}{\sqrt{(2\lambda)}}\frac{(2\nu)!}{2^{\nu}\nu!}\left(\frac{1}{2\lambda} - \right.$$

The appropriate coefficient of H_{2r} in the Type A series is then $\dfrac{\left(\frac{1}{2\lambda} - 1\right)}{2^{\nu}\nu!}$. Now when $x = 0$, $H_{2\nu} = \dfrac{(-1)^{\nu}(2\nu)!}{2^{\nu}\nu!}$. The series then becomes

$$\sum_{\nu=0}^{\infty} \frac{1}{\sqrt{(2\lambda)}}\frac{(2\nu)!}{2^{2\nu}(\nu!)^2}\left(1 - \frac{1}{2\lambda}\right)^{\nu}.$$

In virtue of the Stirling approximation to the factorial, the νth term of this, say u_{ν}, becomes in the limit

$$u_{\nu} \sim \left(1 - \frac{1}{2\lambda}\right)^{\nu}\frac{1}{\sqrt{(2\lambda\pi\nu)}}$$

so that

$$u_{\nu}^{\frac{1}{\nu}} \quad 1 - \frac{1}{2\lambda}.$$

Hence, for $\lambda < \frac{1}{4}$ the series is divergent.

6.29. From the statistical viewpoint, however, the important question is not whether an *infinite* series can represent a frequency function, but whether a finite number of terms

can do so to a satisfactory approximation. It is possible that even when the infinite series diverges its first few terms will give an approximation of an asymptotic character.

This subject has not yet been fully explored and there has been some controversy about the value of the finite Type A series. Two things seem clear :—

(*a*) The sum of a finite number of terms of the series may give negative frequencies, particularly near the tails (as, for instance, in Example 6.2).

(*b*) The series in the Charlier form (6.40) may behave irregularly in the sense that the sum of k terms may give a worse fit than the sum of $(k - 1)$ terms.

How serious these disadvantages may be depends on the purposes in view. So far as practical graduation is concerned it would appear that the finite Type A series is successful only in cases of moderate skewness and in many such cases a Pearson distribution is just as good. In many statistical inquiries we are more interested in the tails of a distribution than its behaviour in the neighbourhood of the mode, and it is here that the Type A series appears particularly inadequate.

But this is not by any means a unanimous view. Arne Fisher (1922) has considered a modified form of the series which he claims to meet most of these criticisms. He considers the series

$$f = (c_0 + c_1 H_1 + \ldots c_r H_r)\alpha(x) \qquad . \quad . \quad . \quad . \quad (6.53)$$

but determines the c's, not from the observed moments and the relations (6.38) but by the method of least squares, i.e. so that

$$\Sigma\{f - (c_0 + c_1 H \ldots + c_r H_r)\alpha(x)\}^2$$

shall be a minimum. The method involves some laborious arithmetic, but Fisher has successfully graduated a number of actuarial experiences by using it.

Two other actuarial statisticians have pointed out the difficulties of the Type A series, Steffensen (1930) adducing some theoretical objections and Elderton (1938) summing up in favour of Pearson distributions.

Example 6.3

As an illustration of the irregular behaviour of terms in the Type A series, consider the distribution

$$dF = \frac{1}{\Gamma(p)} x^{p-1} e^{-x}\, dx \qquad 0 \leqslant x \leqslant \infty.$$

Its characteristic function is

$$(1 - it)^p$$

and thus

$$\kappa_r = p(r - 1)!$$

or, in standard measure,

$$\kappa_r = \frac{(r-1)!}{p^{\frac{r}{2}-1}}.$$

From the manner of the formation of terms in (6.48) it is evident that the coefficient c_r is the sum of terms κ_r, $\kappa_{r-3}\kappa_3$, . . . $(\kappa_{q_1}\kappa_{q_2} \ldots \kappa_{q_m})$, where $(q_1 \ldots q_m)$ is a partition of r such that no q is less than 3. It will then be clear that, since κ_q is of order $p^{1-\frac{q}{2}}$, the term of greatest order in p is that with the greatest number of parts in $(\kappa_{q_1} \ldots \kappa_{q_m})$. For example, if $r = 9$ it is (3^3), if $r = 8$ it is (4^2), and so on.

From these considerations we can find the order in p of the terms in the Type A series. They are

Term	c_0	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
Order in p	0	$-\frac{1}{2}$	-1	$-1\frac{1}{2}$	-1	$-1\frac{1}{2}$	-2	$-1\frac{1}{2}$	-2	$-2\frac{1}{2}$	-2

The terms decrease in order of p, but not at all regularly, and it is clear that in general no coefficient will be negligible compared with a preceding one if p is large. The asymptotic qualities of such series obviously require careful investigation in particular cases.

The Type B Series

6.30. Just as the Type A is derived from the normal integral, a Type B series has been derived by Charlier from the Poisson distribution. Writing

$$\gamma(m, x) = \frac{e^{-m} m^x}{x!} \tag{6.54}$$

for integral values of x, put

$$\gamma(m, x) = \frac{e^{-m}}{\pi}\int_0^{\pi} e^{m \cos t} \cos (m \sin t - xt)\, dt \tag{6.55}$$

for all x. When x is integral this reduces to (6.54).* In other cases

$$\gamma(m, x) = \frac{e^{-m}}{\pi} \sin \pi x \sum_{j=1}^{\infty} (-1)^j \frac{m^j}{j!(x-j)}. \tag{6.56}$$

Write

$$\nabla \gamma = \gamma(m, x) - \gamma(m, x-1)$$

and

$$f(x) = \sum_{j=0}^{\infty} b_j \nabla^j \gamma \tag{6.57}$$

This is the Type B. Charlier recommends it in cases of skewness when Type A is inapplicable (though the dividing-line is not clear). In theory it may be used for continuous variates, but in practice has only been applied to discontinuous variates proceeding by equal intervals. In fact, the objections to Type A apply *a fortiori* to the continuous form of Type B and various other complications appear (cf. Steffensen, 1930).

6.31. Defining polynomials G_r by the relation

$$\gamma(m, x) G_r(m, x) = \frac{d^r}{dm^r} \gamma(m, x) \tag{6.58}$$

we find that

$$G_r = \text{coefficient of } \frac{t^r}{r!} \text{ in } e^{-t}\left(1 + \frac{t}{m}\right)^x$$

$$= \frac{r!}{m^r} \sum_{j=0}^{r} \left\{(-1)^j \binom{x}{r-j} \frac{m^j}{j!}\right\} \tag{6.59}$$

* The integral $\int_0^{2\pi} e^{m e^{it} - ixt} dt$, by the substitution $e^{it} = z$, is 2π times the residue of $\frac{e}{iz^{x+1}}$ in the unit circle and is thus equal to $2\pi \frac{m^x}{x!}$

In a similar manner to that used for the Tchebycheff-Hermite polynomials we have

$$\nabla G_r = \frac{r}{m} G_{r-1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.60)$$

which may be compared with (6.31). It may also be shown that

$$G_r = (-1)^r \frac{\nabla^r \gamma(m, x-r)}{\gamma(m, x)} \quad . \quad . \quad . \quad . \quad . \quad (6.61)$$

and thus G_r may be calculated from the rth differences of the Poisson function $\gamma(m, x)$ in the same way that H_r may be derived from the rth differential coefficient of the normal distribution.

The G's also obey the orthogonal law

$$\left.\begin{aligned}\sum_{x=0}^{\infty}(G_r G_s \gamma) &= 0, \quad r \neq s \\ &= \frac{r!}{m^r}, \ r = s\end{aligned}\right\} \quad . \quad (6.62)$$

Thus if

$$f = \sum_j (b_j G_j \gamma)$$

$$b_j = \frac{m^j}{j!} \sum_{x=0}^{\infty} (f G_j). \quad . \quad . \quad (6.63)$$

TABLE 6.2

Type B Series fitted to a Discontinuous Distribution of Particles emitted by a Radioactive Element in Units of Time.

	Frequency.	Type B (2 terms).	Type B (3 terms).	Type B (4 terms).
0	57	49·5	49·0	58·2
1	203	201·3	201·0	199·8
2	383	403·4	404·3	386·1
3	525	532·3	533·8	523·9
4	532	520·6	521·5	532·1
5	408	402·6	402·5	418·2
6	273	254·8	254·4	260·2
7	139	137·1	136·7	134·0
8	45	64·0	63·9	56·7
9	27	26·1	26·2	22·9
10	10	9·4	9·6	8·6
11	4	3·0	3·1	3·6
12	0	0·9	0·9	1·6
13	1	0·2	0·2	0·8
14	1	0·0	0·0	0·3
TOTALS	2608	2605·2	2607·1	2609·0

In the same manner as for Type A we have, choosing the constant m equal to μ_1',

$$\begin{aligned}
b_0 &= 1 \\
b_1 &= 0 \\
b_2 &= \tfrac{1}{2}(\mu_2 - m) \\
b_3 &= \frac{1}{3!}(\mu_3 - 3\mu_2 + 2m) \\
b_4 &= \frac{1}{4!}\{\mu_4 - 6\mu_3 + \mu_2(11 - 6m) + 3m(m-2)\} \\
&\text{etc.}
\end{aligned} \qquad (6.64)$$

Example 6.4

Table 6.2 shows the frequency of the number of alpha-particles (x) emitted by a bar of polonium in intervals of $\frac{1}{8}$th of a minute in some experiments by Rutherford and Geiger, together with the frequencies given by the Type B series with two, three and four terms. The calculations are due partly to A. Fisher and partly to Aroian (1938).

The Normalisation of Frequency Functions

6.32. Several of the important theoretical distributions occurring in statistics depend on some parameter n in such a way that as n tends to infinity the distribution tends to normality. For large n it is often a sufficient approximation to assume the distribution normal, but for small or moderate n this may be hardly exact enough. In such a case we are nevertheless able to use the normal integral by seeking for a variate transformation

$$\xi = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \qquad (6.65)$$

where the a's are of order $n^{-\frac{1}{2}}$ or smaller. By choosing the a's appropriately we can bring the distribution of ξ much nearer to normality than that of x and hence find the distribution function of x from that of ξ, assumed normal.

Consider in fact the Edgeworth form of the Type A expansion (6.45)

$$\exp\left\{-\frac{\kappa_1 - m}{1!}D + \frac{\kappa_2 - \sigma^2}{2!}D^2 - \frac{\kappa_3}{3!}D^3 + \dots\right\}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(\xi - m)^2} \qquad (6.66)$$

We have retained the terms in D and D^2 because the approximation may perhaps be slightly improved by taking m and σ^2 in the ξ-distribution not quite equal to the mean and variance of x.

We now assume that the cumulant κ_r is of order n^{1-r}, a case of fairly common occurrence; that $\kappa_1 - m$ is, by choice of m, of order n^{-1}; and that $\kappa_2 - \sigma^2$, by choice of σ^2, is of order n^{-2}, so that we may write

$$\begin{aligned}
\kappa_1 - m &= l_1\sigma \qquad & l_1 &= 0(n^{-\frac{1}{2}}) \\
\kappa_2 - \sigma^2 &= l_2\sigma^2 \qquad & l_2 &= 0(n^{-1})
\end{aligned}$$

Then σ^2 is of order κ_2, i.e. n^{-1}, and thus

$$\frac{\kappa_r}{\sigma^r} = 0(n^{1-\frac{1}{2}r}).$$

Thus (6.66) may be written

$$\exp\{-l_1\sigma D + \tfrac{1}{2}l_2\sigma^2 D^2 - \tfrac{1}{6}l_3\sigma^3 D^3 + \tfrac{1}{24}l_4\sigma^4 D^4 - \tfrac{1}{120}l_5\sigma^5 D^5 + \tfrac{1}{720}l_6\sigma^6 D^6 - \dots\}\frac{1}{\sqrt{(2\pi)}\sigma}e^{-\frac{1}{2\sigma^2}(x-m)^2} \qquad (6.67)$$

where l_1 and l_3 are $0(n^{-\frac{1}{2}})$
l_2 and l_4 are $0(n^{-1})$
l_5 is $0(n^{-\frac{3}{2}})$
l_6 is $0(n^{-2})$, etc.

Expanding the operator and retaining only terms up to and including $0(n^{-2})$ we find for the operator

$$\begin{aligned}&1 - l_1\sigma D + \tfrac{1}{2}l_2\sigma^2D^2 - \tfrac{1}{6}l_3\sigma^3D^3 + \tfrac{1}{24}l_4\sigma^4D^4 - \tfrac{1}{120}l_5\sigma^5D^5 + \tfrac{1}{720}l_6\sigma^6D^6 + \tfrac{1}{2}(l_1^2\sigma^2D^2\\&+\tfrac{1}{4}l_2^2\sigma^4D^4 + \tfrac{1}{36}l_3^2\sigma^6D^6 + \tfrac{1}{576}l_4^2\sigma^8D^8 - l_1l_2\sigma^3D^3 + \tfrac{1}{3}l_1l_3\sigma^4D^4 - \tfrac{1}{12}l_1l_4\sigma^5D^5\\&+\tfrac{1}{60}l_1l_5\sigma^6D^6 - \tfrac{1}{6}l_2l_3\sigma^5D^5 + \tfrac{1}{24}l_2l_4\sigma^6D^6 - \tfrac{1}{72}l_3l_4\sigma^7D^7 + \tfrac{1}{360}l_3l_5\sigma^8D^8) + \tfrac{1}{6}(-l_1^3\sigma^3D^3\\&-\tfrac{1}{216}l_3^3\sigma^9D^9 + \tfrac{3}{2}l_1^2l_2\sigma^4D^4 - \tfrac{1}{2}l_1^2l_3\sigma^5D^5 - \tfrac{1}{8}l_1^2l_4\sigma^6D^6 + \tfrac{1}{288}l_3^2l_4\sigma^{10}D^{10} - \tfrac{1}{12}l_1l_3^2\sigma^7D^7\\&+\tfrac{1}{24}l_2l_3^2\sigma^8D^8 + \tfrac{1}{2}l_1l_2l_3\sigma^6D^6 + \tfrac{1}{24}l_1l_3l_4\sigma^8D^8) + \tfrac{1}{24}(l_1^4\sigma^4D^4 + \tfrac{1}{1296}l_3^4\sigma^{12}D^{12}\\&+\tfrac{2}{3}l_1^3l_3\sigma^6D^6 + \tfrac{1}{6}l_1^2l_3^2\sigma^8D^8 + \tfrac{1}{54}l_1l_3^3\sigma^{10}D^{10}).\quad . \quad . \quad . \quad (6.68)\end{aligned}$$

The result of this operation is a similar expression, which we will not bother to write out at length, with the operator σ^rD^r replaced by $(-1)^rH_r\left(\dfrac{xm}{\sigma}\right)$ and multiplied by $\dfrac{1}{\sigma\sqrt{(2\pi)}}$ $e^{-\frac{1}{2\sigma^2}(x-m)^2}$.

The distribution function is given by integrating this expression, and we then have for the frequency less than or equal to $m + \sigma x$ (arranging the terms in order of magnitude in n)

$$\begin{aligned}&\int_{-\infty}^{x}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}\,dx + \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}[-(l_1 + \tfrac{1}{6}l_3H_2) - (\tfrac{1}{2}l_1^2H_1 + \tfrac{1}{2}l_2H_1 + \tfrac{1}{6}l_1l_3H_3\\&+\tfrac{1}{24}l_4H_3 + \tfrac{1}{72}l_3^2H_5) - (\tfrac{1}{6}l_1^3H_2 + \tfrac{1}{2}l_1l_2H_2 + \tfrac{1}{12}l_1^2l_3H_4 + \tfrac{1}{12}l_2l_3H_4 + \tfrac{1}{24}l_1l_4H_4\\&+\tfrac{1}{120}l_5H_4 + \tfrac{1}{72}l_1l_3^2H_6 + \tfrac{1}{144}l_3l_4H_6 + \tfrac{1}{1296}l_3^3H_8) - (\tfrac{1}{24}l_1^4H_3 + \tfrac{1}{8}l_2^2H_3 + \tfrac{1}{4}l_1^2l_2H_3\\&+\tfrac{1}{36}l_1^3l_3H_5 + \tfrac{1}{12}l_1l_2l_3H_5 + \tfrac{1}{48}l_1^2l_4H_5 + \tfrac{1}{48}l_2l_4H_5 + \tfrac{1}{120}l_1l_5H_5 + \tfrac{1}{720}l_6H_5\\&+\tfrac{1}{144}l_1^2l_3^2H_7 + \tfrac{1}{144}l_2l_3^2H_7 + \tfrac{1}{1152}l_4^2H_7 + \tfrac{1}{144}l_1l_3l_4H_7 + \tfrac{1}{720}l_3l_5H_7 - \tfrac{1}{1296}l_1l_3^3H_9\\&+\tfrac{1}{1728}l_3^2l_4H_9 + \tfrac{1}{31104}l_3^4H_{11})]\quad . \quad . \quad . \quad (6.69)\end{aligned}$$

6.33. Now let ξ be a normal variate. We will determine ξ in terms of x such that

$$\int_{-\infty}^{\xi}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{y^2}{2}}\,dy = F(x), \quad . \quad (6.70)$$

$F(x)$ being the distribution function given by the Type A expansion (6.69).

We have

$$\begin{aligned}\int_{-\infty}^{\xi}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{y^2}{2}}\,dy &= F(\xi) = F(x + \xi - x)\\&= \int_{-\infty}^{x}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{y^2}{2}}\,dy - \frac{(x-\xi)}{1!}\frac{d}{dx}\int_{-\infty}^{x}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{y^2}{2}}\,dy \quad . \; . \; . \text{ etc.,}\end{aligned}$$

by Taylor's theorem,

$$= \int_{-\infty}^{x}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{y^2}{2}}\,dy - \frac{(x-\xi)}{1!}\alpha(x) \quad \frac{(x-\xi)^2}{2!}H_1(x)\alpha(x) - \frac{(x-\xi)^3}{3!}H_2(x)\alpha(x) \quad . \quad (6.71)$$

and this is equal to (6.69).

The next step is to invert this series so as to give $(x - \xi)$ in powers of x. Assuming

$$x - \xi = a_0 + a_1x + a_2x^2 + \ . \ . \ . \text{ etc.,} \quad . \quad . \quad . \quad (6.72)$$

we see that when $x = 0$, $\xi = -a_0$, $x - \xi$ is of order $n^{-\frac{1}{2}}$; and hence, to order n^{-2} we have from (6.71), with $x = 0$,

$$\xi - \frac{\xi^2}{2}(+0) + \frac{\xi^3}{6}(-1) = a_0 - \frac{{a_0}^3}{6}$$

and this is equal to the expression in square brackets in (6.69) with $x = 0$.

We then find

$$a_0 = l_1 - \tfrac{1}{6}l_3 - \tfrac{1}{2}l_1l_2 + \tfrac{1}{6}l_1^2l_3 + \tfrac{1}{4}l_2l_3 + \tfrac{1}{8}l_1l_2 + \tfrac{1}{40}l_5 - \tfrac{7}{36}l_1l_3^2 - \tfrac{15}{144}l_3l_4 + \tfrac{52}{648}l_2l_4.$$

We can now find a_1 in (6.72) by identifying coefficients in x, and so on. After some algebraic reduction we find, writing the terms in descending order in n,

$$\begin{aligned} x - \xi = l_1 &+ \tfrac{1}{6}l_3(x^2 - 1) + \tfrac{1}{2}l_2x - \tfrac{1}{3}l_1l_3x + \tfrac{1}{24}l_4(x^3 - 3x) - \tfrac{1}{36}l_3^2(4x^3 - 7x) - \tfrac{1}{2}l_1l_2 \\ &+ \tfrac{1}{6}l_1^2l_3 - \tfrac{1}{12}l_2l_3(5x^2 - 3) - \tfrac{1}{8}l_1l_2(x^2 - 1) + \tfrac{1}{120}l_5(x^4 - 6x^2 + 3) + \tfrac{1}{36}l_1l_3^2(12x^2 - 7) \\ &- \tfrac{1}{144}l_3l_4(11x^4 - 42x^2 + 15) + \tfrac{1}{648}l_3^3(69x^4 - 187x + 52) - \tfrac{3}{8}l_2^2x + \tfrac{5}{6}l_1l_2l_3x \\ &+ \tfrac{1}{8}l_1^2l_4x - \tfrac{1}{48}l_2l_4(7x^3 - 15x) - \tfrac{1}{30}l_1l_5(x^3 - x) + \tfrac{1}{720}l_6(x^5 - 10x^3 + 15x) \\ &- \tfrac{1}{3}l_1^2l_3^2x + \tfrac{1}{72}l_2l_3^2(36x^3 - 49x) - \tfrac{1}{384}l_4^2(5x^5 - 32x^3 + 35x) + \tfrac{1}{36}l_1l_3l_4(11x^3 - 21x) \\ &- \tfrac{1}{360}l_3l_5(7x^5 - 48x^3 + 51x) - \tfrac{1}{324}l_1l_3^3(138x^3 - 187x) + \tfrac{1}{864}l_3^2l_4(111x^5 \\ &- 547x^3 + 456x) - \tfrac{1}{7776}l_3^4(948x^5 - 3628x^3 + 2473x). \end{aligned} \tag{6.73}$$

This is our required expression of the variate ξ in terms of the variate x. To order n^{-2} at least ξ will be normally distributed.

It is often more convenient to express x in terms of ξ. This may be done by noting that

$$\begin{aligned} x - \xi = g(x) &= g(\xi + x - \xi) \\ &= g(\xi) + (x - \xi)g'(\xi) + \ldots \\ &= g(\xi) + g'(\xi)\{g(\xi) + x - \xi g'(\xi)\} + \ldots \end{aligned}$$

and by continuing the process

$$\begin{aligned} x - \xi = g(\xi) &+ g(\xi)g'(\xi) + g(\xi)g'^2(\xi) + \tfrac{1}{2}g^2(\xi)g''(\xi) + g(\xi)g'^3(\xi) + \tfrac{3}{2}g^2(\xi)g'(\xi)g''(\xi) \\ &+ \tfrac{1}{6}g^3(\xi)g'''(\xi) + \ldots \end{aligned} \tag{6.74}$$

Hence, using the value of ξ given by (6.73) we find, after some reduction,

$$\begin{aligned} x - \xi = l_1 &+ \tfrac{1}{6}l_3(\xi^2 - 1) + \tfrac{1}{2}l_2\xi + \tfrac{1}{24}l_4(\xi^3 - 3\xi) - \tfrac{1}{36}l_2^2(2\xi^3 - 5\xi) - \tfrac{1}{6}l_2l_3(\xi^2 - 1) \\ &+ \tfrac{1}{120}l_5(\xi^4 - 6\xi^2 + 3) - \tfrac{1}{24}l_3l_4(\xi^4 - 5\xi^2 + 2) + \tfrac{1}{324}l_3^3(12\xi^4 - 53\xi^2 + 17) \\ &- \tfrac{1}{8}l_2^2\xi - \tfrac{1}{16}l_2l_4(\xi^3 - 3\xi) + \tfrac{1}{720}l_6(\xi^5 - 10\xi^3 + 15\xi) + \tfrac{1}{72}l_2l_3^2(10\xi^3 - 25\xi) \\ &- \tfrac{1}{384}l_4^2(3\xi^5 - 24\xi^3 + 29\xi) - \tfrac{1}{180}l_3l_5(2\xi^5 - 17\xi^3 + 21\xi) + \tfrac{1}{288}l_3^2l_4(14\xi^5 \\ &- 103\xi^3 + 107\xi) - \tfrac{1}{7776}l_3^4(252\xi^5 - 1688\xi^3 + 1511\xi) \end{aligned} \tag{6.75}$$

Example 6.5

Consider again the distribution of Example 6.3—

$$dF = \frac{1}{\Gamma(p)}e^{-x}x^{p-1}\,dx \qquad 0 \leqslant x \leqslant \infty.$$

We have already found that, in standard measure, this tends to the normal form, and that κ_r is of order $p^{1-\frac{r}{2}}$.

We will take l_1 and l_2 of (6.67) to be zero, which implies that our normal variate ξ is to have the same mean and variance as that of x. We have

$$l_3 = 2p^{-\frac{1}{2}} \qquad l_4 = 6p^{-1} \qquad l_5 = 24p^{-\frac{3}{2}} \qquad l_6 = 120p^{-2}.$$

(6.69) then becomes

$$\int_{-\infty}^{\cdot}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}\,dx-\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}\left\{\frac{1}{3p^{\frac{1}{2}}}H_2+\frac{1}{4p}H_3+\frac{1}{18p}H_5+\frac{1}{5p^{\frac{3}{2}}}H_4+{}_{12p}\mathrm{H}_6+{}_{162p}\mathrm{H}_8\right.$$
$$\left.+\frac{1}{6p^2}H_5+\frac{1}{32p^2}H_7+\frac{1}{15p^2}H_7+\frac{1}{72p^2}H_9+\frac{1}{1944p^2}H_1\right.$$

Let us, as a simple illustration, find the distribution function of x for $p = 9$, $x = 12$. The mean of the distribution is then 9 and its variance 9, so that this corresponds to a deviation $(12-9)\sqrt{9}$ in standard measure, equal to unity. It is found from (6.30) and an additional equation for H_{11} that

$$\mathrm{H}_2 = 0,\ H_3 = -2,\ H_4 = -2,\ H_5 = 6,\ H_6 = 16,\ H_7 = -20,\ H_8 = -132,\ H_9 = 28,$$
$$H_{10} = 1216,\ H_{11} = 936.$$

We then find for the distribution function

$$\int_{-\infty}^{\cdot}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx+\frac{1}{\sqrt{(2\pi)}}e^{-\frac{x^2}{2}}(0{\cdot}015{,}163{,}5).$$

The values for the normal function are obtained from the tables and we get the value

$$0{\cdot}841{,}345 + (0{\cdot}241{,}970{,}7)(0{\cdot}015{,}163{,}5) = 0{\cdot}8450,$$

which is exact to four places. The approximation is evidently fairly good, even for values of p as low as 9.

We could have found the same result by using (6.73). Substituting $x = 1$ in that equation we find

$$\xi = 1{\cdot}015{,}386,$$

and the distribution function for the normal integral with deviate equal to this value of ξ is 0·8450 as before.

Suppose now we wish to find the deviate x whose distribution function is $F(x) = 0{\cdot}99$ when $p = 15$.

The normal deviate ξ corresponding to such a value is found from tables to be 2·326,348. We then have from (6·75)

$$x-\xi=\frac{1}{3p^{\frac{1}{2}}}(\xi^2-1)+\frac{1}{4p}(\xi^2-3\xi)+,\text{ etc.},$$

which will be found to give

$$x = 2{\cdot}697{,}22.$$

This is the value in standard measure. The deviate in ordinary measure is

$$15 + x\sqrt{15} = 25{\cdot}45.$$

This is exact to two places of decimals.

The example shows that, notwithstanding the non-convergence of the infinite Type A series, a satisfactory approximation may be obtained from its first few terms, at least in certain cases. We may remark without proof that by an adaptation of a procedure given by Cramér (1928) it may be shown that an asymptotic expansion does in fact exist for the distribution of this example.

NOTES AND REFERENCES

An excellent account of Pearson's distributions is given in Elderton's book. Examples of the fitting of the distributions to the data of experience abound in *Biometrika*.

For the Type A series see Charlier (1906 and 1931), Henderson (1922), Cramér (1928) and Bowley (1928). For the Type B series see Charlier, Jordan (1927), Aroian (1937) and Steffensen (1930).

Charlier has also proposed a Type C series, as to which see his paper of 1928 and the brochure of 1931.

For the convergence of the Type A series and its relationship to elementary errors see two admirable papers by Cramér (1926 and 1928).

A very good general account of these distributions and an examination of the possibility of extending them to the bivariate case is given by Pretorius (1930), who gives a number of references. Up to the present no entirely satisfactory system of bivariate distributions corresponding either to those of Pearson or to those of Charlier has been found.

For some early efforts by Edgeworth to transform distributions to the normal form, see Bowley (1928) and Pretorius (1930). The approach of sections **6.32** and **6.33** is due to Cornish and Fisher (1937), who give some tables which are useful in this type of work.

The polynomials $H_r(x)$ are frequently referred to by English writers as Hermite polynomials, but they are really due to Tchebycheff (*Mémoires de l'Academie de Saint Pétersbourg*, 1860). Hermite's papers on this subject followed four years later (*Comptes rendus*, 58, 93 and 266).

Aroian, L. A. (1937), "The Type B Gram-Charlier series," *Ann. Math. Statist.*, **8**, 183.

Bowley, A. L. (1928), *F. Y. Edgeworth's Contributions to Mathematical Statistics*, Royal Statistical Society.

Charlier, C. V. L. (1906), *Researches into the Theory of Probability*, Lund.

—— (1928), "A new form of the frequency function," *Meddelande fröm Lunds Astronomiska Observatorium*, Series II, No. 51.

—— (1931), *Applications à l'astronomie* (one of the series in Borel's *Traité du calcul des Probabilités*, Gauthier-Villars, Paris).

Comrie, L. J. (1939), *Tables of* $tan^{-1}x$ *and log* $(1 + x^2)$, Cambridge University Press.

Cornish, E. A., and Fisher R. A. (1937), "Moments and cumulants in the specification of distributions," *Revue de l'Inst. Int. Stat.*, **5**, 307.

Cramér, H. (1926), "On some classes of series used in mathematical statistics," *Den sjette Skandinaviske Matematikercongres*, Copenhagen.

—— (1928), "On the composition of elementary errors," *Skandinavisk Aktuarietidskrift*, 13 and 141.

Edgeworth, F. Y. (1904), "The Law of Error," *Cambridge Phil. Trans.*, **20**, 36, 113 (and an appendix issued with bound reprints).

Elderton, Sir W. P. (1938), *Frequency Curves and Correlation*, 3rd edition, Cambridge University Press.

Fisher, Arne (1922), *Frequency Curves*, Macmillan.

Henderson, J. (1922), "On expansions in tetrachoric functions," *Biometrika*, **14**, 157.

Jordan, C. (1927), *Statistique Mathématique*, Gauthier-Villars, Paris.

Jørgensen, N. R. (1916), *Undersøgelser over Frequensflader og Korrelation*, Busck, Copenhagen.

Pearson, K. (1925), "The fifteen constant bivariate frequency surface," *Biometrika*, **17**, 268.

Pretorius, S. J. (1930), "Skew bivariate frequency surfaces examined in the light of numerical illustrations," *Biometrika*, **22**, 109.

Romanovsky, V. (1924), "Generalisation of some types of the frequency curves of Professor Pearson," *Biometrika*, **16**, 106.

Steffensen, J. (1930), *Some Recent Researches in the Theory of Statistics and Actuarial Science*, Cambridge University Press.
Wishart, J. (1926), "On Romanovsky's generalised frequency curves," *Biometrika*, **18**, 221.

EXERCISES

6.1. Show that for the Pearson distributions

$$\frac{d \log y}{dx} = \frac{x}{B_0 + B_1 x + B_2 x^2}$$

the range is unlimited in both directions if $B_0 + B_1 x + B_2 x^2$ has no real roots; limited in one direction if the roots are real and of the same sign; and limited in both directions if the roots are real and of opposite sign.

6.2. Show that the Pearson Type VI curve may be written

$$y = y_0\left(1 - \frac{x^2}{a^2}\right)^{-m} e^{-\nu \tanh^{-1} \frac{x}{a}}$$

and discuss the relationship with the Type IV curve.

6.3. Assign the following distributions to one of Pearson's types:—

$$dF = ke^{-\frac{\chi^2}{2}}\chi^{\nu-1}\, d\chi$$

$$dF = \frac{k\, dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

$$dF = k(1 - r^2)\, dr$$

$$dF = k\eta^{p-3}(1 - \eta^2)^{\frac{N-p-2}{2}}\, d\eta$$

(All these distributions are important in the theory of sampling.)

6.4. Show that for the Type B series the coefficients of equation (6.63) may be written symbolically—

$$b_j = \frac{1}{j!}(\mu' - m)^{[j]}$$

$$= \frac{1}{j!}\left\{\mu'_{[j]} - \binom{j}{1}\mu'_{[j-1]}m + \binom{j}{2}\mu'_{[j-2]}m^2 \ldots\right\}$$

(C. Jordan, 1927.)

6.5. Show that, in the notation of **6.30**,

$$\nabla\gamma(m, x) = -\frac{d}{dm}\gamma(m, x + 1).$$

Hence that

$$\sum_{x=0}^{\lambda}\gamma(m, x) = 1 - I_m(\lambda + 1).$$

where I_m is the incomplete Γ-function; and hence that the sum of the first $(\lambda + 1)$ terms of the Type B series is given by

$$b_0\{1 - I_m(\lambda + 1)\} - (b_1 + b_2 G_1 + \ldots)\gamma(m, \lambda)$$

(C. Jordan, 1927.)

6.6. Show that if y is a function of x which it is desired to represent approximately by the form

$$y = \sum_{j=0}^{r} c_j H_j(x)\alpha(x),$$

then the values of the c's appropriate to the expansion of y in this form are such as to minimise the sum

$$\int_{-\infty}^{\infty}\left\{\frac{y - \sum_{j=0} c_j H_j(x)\alpha(x)}{\sqrt{\alpha(x)}}\right\}^2 dx$$

6.7. Show that for a Pearson distribution $\dfrac{df}{f} = \dfrac{a + x}{b_0 + b_1 x + b_2 x^2}$ the characteristic function obeys the relation

$$b_2\theta\frac{d^2\phi}{d\theta^2} + (1 + 2b_2 + b_1\theta)\frac{d\phi}{d\theta} + (a + b_1 + b_0\theta)\phi = 0,$$

where $\theta = it$. Deduce the recurrence relation between the moments.

Show also that the cumulative function obeys the relation

$$b_2\theta\left\{\frac{d^2\psi}{d\theta^2} + \left(\frac{d\psi}{d\theta}\right)^2\right\} + (1 + 2b_2 + b_1\theta)\frac{d\psi}{d\theta} + (a + b_1 + b_0\theta) = 0.$$

Hence show that the cumulants obey the recurrence relations

$$\{1 + (r + 2)b_2\}\kappa_{r+1} + rb_1\kappa_r + rb_2\left\{\binom{r-1}{1}\kappa_2\kappa_{r-1} + \binom{r-1}{2}\kappa_3\kappa_{r-2} + \ldots \right.$$
$$\left. + \binom{r-1}{j}\kappa_{j+1}\kappa_{r-j} + \ldots + \binom{r-1}{1}\kappa_{r-1}\kappa_2\right\} = 0.$$

6.8. Show that no distribution which is not completely determined by its moments can be expanded in a convergent Type A series.

6.9. If the distribution

$$dF = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}\,dx \qquad -\infty \leqslant x \leqslant \infty$$

is transformed by

$$x = \frac{1}{k}(\log_{10}\xi - l)$$

and

$$\frac{1}{b} = \log_{10}e, \qquad \lambda = e^{b^2k^2},$$

show that

$$\beta_1 = \lambda^2(\lambda + 3) - 4$$
$$\beta_2 - 3 = \lambda^2(\lambda^2 + 2\lambda + 3) - 6$$

and that

$$\mu_3(\mu_1')^3 - 3\mu_2^2(\mu_1')^2 - \mu_2^3 = 0,$$

where μ_1' is the first moment about the start of the transformed curve. Thence that

$$l = 2 \log \mu_1' - \tfrac{1}{2} \log (\mu_2 + \mu_1'^2)$$
$$bk^2 = 2 \log \mu_1' - 2l.$$

6.10. Show that if a function in standard measure is expanded in a Type A series the coefficients of the second and third terms depend respectively on β_1 and β_2 and thus provide measures of skewness and kurtosis.

CHAPTER 7

PROBABILITY AND LIKELIHOOD

7.1. The previous six chapters have dealt with the theory of statistical distributions from a descriptive point of view. It has been explained that the distributions occurring in practice exhibit certain regular features which permit of representation by mathematical forms; that they can be characterised by certain parameters such as moments and cumulants; and that certain general theorems about distribution and frequency functions can be deduced. We now begin a study of a different kind, namely, the inquiry whether any statements can be made about populations or their parameters and distributions when only a sample of the populations is available for scrutiny. Except in trivial cases it is not possible to make any statements on these matters with the categorical certainty of deductive logic; but it is possible, and indeed it is necessary if scientific inquiry is to go forward at all, to make statements of a less definite nature in terms of probability. In this chapter we shall accordingly be concerned with the theory of probability as it affects statistics and in subsequent chapters with its applications in statistical theory.

7.2. In ordinary speech we use the words "probability", "chance" or "likelihood" to describe an attitude of mind towards some proposition of whose truth we are not certain. We say that it is improbable that life exists on Mars, that the chances are that if a penny is tossed ten times it will come down heads at least once, that it is likely to rain to-morrow, and so on. It is rarely indeed in practical affairs that we are confronted with a proposition of whose truth we are absolutely certain. Nevertheless, we often have to assume that such propositions are true or untrue in order to reach decisions and to act in a rational way.

The attitude of doubt we adopt is described in terms of probability. We say that the propositions are more or less probable and accept or reject them accordingly.

7.3. A little introspection will convince the reader that all the attitudes of mind to which we relate the concept of probability have certain things in common:—

(*a*) They concern *propositions*. The mind considers a proposition which has meaning and assumes towards it a certain attitude of doubt. It is very common both in mathematics and in statistics to speak of the probability of an event, or even of a variate-value; but these are condensed expressions for the proposition that an event will happen or that a member of a population has a given variate-value, and, though very convenient shorthand expressions which will often be used in the sequel, must not be allowed to obscure the essential fact that propositions are concerned.

(*b*) There are *degrees* of probability. We say that it is very improbable that a hundred tosses with a penny will not result in a head; that it is more probable that horse A will win a race than that horse B will do so; that the probability of having wet weather in the course of an English summer is so great as to be near certainty. But it does not follow (and some writers on the logic of probability do not admit) that every pair of probabilities can be compared. It could with consistence be maintained that, whereas we may compare the probability of getting ten trumps in a game of cards with that of getting eight, there is no way of comparing the probabilities of the propositions, say, that there exists a planet outside the orbit of Pluto and that the human race will ultimately go bald.

(c) The degree of probability attributed to a proposition varies according to the amount of relevant evidence available to the particular mind considering the proposition. If we know that a horse has won its three previous races we attach a greater probability to the proposition that it will win the next. If we know that a penny has heads on both sides the probability that it will come down heads when tossed is so great as to amount to certainty; and so on.

(d) Pursuing this last point, we see that certainty can be regarded as a limiting form of probability. As a proposition becomes more and more probable it tends towards certain truth; as it becomes more and more improbable it tends towards certain untruth.

7.4. The object of the theory of probability is to give to the somewhat indefinite notions described above the precision of a science, and, since numerical measurement is the greatest precision which a science can possess, to measure probability numerically. Several writers have explored the more general problem, foreshadowed as early as Leibniz, of developing a logic of probabilities, and the reader who is interested may refer to the work of Keynes (1921), F. P. Ramsey (1931) and Johnson (1921–4). From the statistical viewpoint the interest of this subject centres in the *numerical* theory of probability which alone will concern us in this book.

It is at this point that we arrive at the first of the differences of opinion among authorities on the theory of probability. Some writers try to include all the ideas generally associated with the word "probability" within the scope of their theory, which is thus applicable to any of the attitudes of doubt covered by the meaning of the word. The principal modern exponent of this viewpoint is Jeffreys, whose book (1939) should certainly be read by all serious students of the subject. Most statisticians, on the other hand, are concerned with the probabilities of propositions of a particular kind, namely, those which form the members of *populations* of propositions. Under the more general theory, it has meaning to speak of the probability of an isolated proposition such as the one that Shakespeare's plays were written by Francis Bacon. In statistics we are more usually concerned with the proposition which asserts the happening of some event which could have arisen in a specified number of ways, such as the throwing of a number with an ordinary die. The first approach takes probability to be an undefined idea, like the straight line of Euclidean geometry, and builds up the theory from certain axioms. The second approach seeks to define probability in terms of the relative frequency of events and thus to throw the theory back on to the pure mathematics of abstract ensembles (Kolmogoroff, 1933) or to the limiting properties of sequences (von Mises, 1936). The reader who is perplexed by the controversy between the adherents of the axiomatic and the frequency theories will find many of his difficulties resolved by the consideration that the two theories cover different domains of thought, or rather, that the axiomatic theory attempts to cover a wider domain than the frequency theory.

7.5. This, however, does not explain away the whole of the difficulty, and the reader will have to choose for himself among the various possible sets of fundamental ideas forming the starting-point of the theory. When we consider the concept of probability as a psychological matter we can either suppose that further analysis is impossible or unprofitable, in which case the axiomatic approach seems inevitable; or we can ask how the mind comes to take up an attitude of belief in propositions which confront it. It is not necessary here to go into this question at length, but there would, in my own opinion, be a considerable measure of agreement that the concept of probability is founded on our experience of the

frequency of observed phenomena. When we say that the probability of a coin coming down heads on being tossed is one-half we have in mind, I think, that if it is tossed a large number of times it will come down heads in approximately half the cases. Even in extreme cases, say, when we attempt to assess the probability of a horse winning a given race, an event which cannot be repeated, we are, I think, picturing our estimation as one of a number of similar acts and assessing the relative frequency of the horse's victory in that population.

But it has to be admitted that, even if this be true, there is no *necessity* to use the concept of frequency in the axiomatisation of the theory. The concept of a straight line may very well be founded on our experience of the local properties of rays of light, but it does not follow that the indefinables of Euclidean geometry are to be analysed into optical concepts.

The Basic Rules of Direct Probability

7.6. For our present purposes the problems of fundamentals may be passed over, since all parties are agreed on the rules governing the calculus of direct probabilities. (The so-called " inverse " probabilities will require more discussion and will be dealt with later.) We therefore enunciate these rules without attempting to deduce them from more primitive propositions.

In the first place it is assumed that probability is measurable on a continuous scale, so that any probability can be expressed as a real number. We shall, in fact, say that a probability *is* x, a real number. This assumption implies, among other things, that any two probabilities may be compared; for if they are measured by the numbers x and y we may say that the probability of the first is greater than, equal to, or less than, that of the second according as $x > y$, $x = y$, or $x < y$.

7.7. The probability of a proposition q on data p is written $P(q \mid p)$. We have then *Rule 1*:—

$$\text{If } p \text{ entails } q,\ P(q \mid p) = 1 \qquad (7.1)$$

$$\text{If } p \text{ entails not-}q,\ P(q \mid p) = 0 \qquad (7.2)$$

This rule defines the end-points of our scale of probability. Certainty that a proposition is not true is represented by zero, certainty that it is true by unity. Any probability lies in the range 0 to 1.

7.8. *Rule 2*.—If $q_1 \ldots q_n$ are a set of equally probable and mutually exclusive propositions on data p, and if Q is a subset of m of these propositions, then

$$P(Q \mid p) = \frac{m}{n}. \qquad (7.3)$$

This proposition is the starting-point of the frequency theory of probability. It is usually stated in some such form as: if of a set of n mutually exclusive and equally probable events m are distinguished by some characteristic A, the probability of an event bearing A is $\frac{m}{n}$

The objection to this rule from the logical viewpoint is that it contains the concept " equally probable " and is thus circular if one adopts it as a definition. The mathematical theorist dealing with probability, in the mathematician's facile way, overcomes this trouble either by accepting the circular definition, or by defining probability purely as a property of sets of points. For example, such a definition might be: if of an aggregate of objects

n in number m are characterised by some quality A, the probability of any member bearing A is, by definition, the number $\frac{m}{n}$. To take a more sophisticated line, we can regard the objects as points of a set, attach set-functions to them obeying certain axioms and postulates, and thus build up the theory of probability as a branch of the theory of set-functions. Any verification of the theory, any test whether it provides a reasonably accurate picture of the way things happen in the world, is referred to experimental physics. The mathematician, of course, is used to this devolution of responsibilities, but the statistician is concerned with concordance between theory and practice and cannot always leave experimental verification to others.

7.9. *Rule 3.*—If the probabilities of n mutually exclusive propositions $q_1 \ldots q_n$ on data p are $P_1 \ldots P_n$, then the probability on data p that one of them is true is $P_1 + P_2 \ldots + P_n$.

This is generally known as the "theorem of the addition of probabilities". In the language of the textbooks, the probability that one of n mutually exclusive events will happen is the sum of their separate probabilities.

7.10. *Rule 4.*—The probability of two propositions q and r on data p is the product of the probability of q given p and that of r given q and p. Symbolically,

$$P(qr \mid p) = P(q \mid p)P(r \mid qp). \qquad (7.4)$$

Since q and r appear symmetrically we also have

$$P(qr \mid p) = P(r \mid p)P(q \mid rp). \qquad (7.5)$$

From the frequency standpoint this rule is almost self-evident. If of a set n, (a) bear the characteristic A, (b) the characteristic B, and (ab) both characteristics, then the rule states that

$$\frac{(ab)}{n} = \frac{(a)}{n}\cdot\frac{(ab)}{(a)} = \frac{(b)}{n}\frac{(ab)}{(b)},$$

a simple arithmetical proposition.

More generally we have

$$P(q_1q_2 \ldots q_k \mid p) = P(q_1 \mid p)P(q_2 \mid q_1p)P(q_3 \mid q_2q_1p) \ldots P(q_k \mid q_{k-1} \ldots q_1p) \quad (7.6)$$

a result which follows from the repeated application of Rule 4.

If, as a particular case,

$$P(qr \mid p) = P(q \mid p)P(r \mid p) \qquad (7.7)$$

we have, in virtue of (7.4),

$$P(r \mid p) = P(r \mid qp), \qquad (7.8)$$

and q is then said to be irrelevant to r, given p. A knowledge of q does not affect the probability of r on data p.

7.11. The above four rules and various elaborations of them provide the basis of the *direct* theory of probability, which is concerned with problems of the type : given a set of propositions with known probabilities, determine the probability of some contingent proposition. This is a branch of pure mathematics and will be found discussed, for example, in most textbooks of algebra. Ultimately all problems in this branch of the theory are reducible to the counting of the number of ways in which certain events can happen. The following examples will illustrate the type of investigation involved.

Example 7.1

What is the probability that a specified player will get a hand containing 13 cards of one suit at a single deal at a game of bridge?

We have to consider here the total number of ways in which a given player can be dealt a hand of cards. There are 52 cards and 13 can be chosen from them in $\binom{52}{13}$ ways. Of these ways only four will contain cards of one suit.

We then assume that all the possible deals are equally probable and are thus able to apply Rule 2. Here $m = 4$ and $n = \binom{52}{13}$, so that the probability is

$$P = \frac{4}{\binom{52}{13}}$$

$$= \frac{4 . 39!\, 13!}{52!}.$$

Factorial expressions of this kind may be found from tabled logarithms of factorials or by the use of the Stirling approximation. In this particular case we find

$$P = 6 \times 10^{-12} \text{ approximately.}$$

Example 7.2

n letters, to each of which corresponds an envelope, are placed in the envelopes at random. What is the probability that no letter is placed in the right envelope?

The condition that the letters are put in the envelopes "at random" is to be interpreted as meaning that every possible way of assigning the letters to envelopes is equally probable. The question, under Rule 2, then reduces to the purely algebraic one: in what proportion of the possible cases does no letter get into the right envelope?

Suppose that u_n is the number of ways in which all the letters go wrong. Consider any two letters. If these occupy each other's envelopes, the number of ways in which the remaining $n - 2$ letters can go wrong is u_{n-2}; and there are $(n - 1)$ ways in which two letters can be interchanged. But if one letter occupies another's place and not vice-versa, which can happen in $(n - 1)$ ways, there are u_{n-1} ways in which the others can go wrong. Hence we have the difference equation

$$u_n = (n - 1)(u_{n-1} + u_{n-2})$$

We may re-write this

$$u_n - nu_{n-1} = -(u_{n-1} - n - 1\ u_{n-2}$$

and putting

$$v_n = u_n - nu_{n-1}$$

we find

$$v_n = - v_{n-1},$$
$$= (-1)^{n-2} v_2.$$

Thus

$$u_n - nu_{n-1} = (-1)^{n-2}(u_2 - u_1).$$

But $u_1 = 0$ and $u_2 = 1$ and thus

$$u_n - nu_{n-1} = (-1)^n$$

whence

$$u_n = n!\left\{\frac{1}{2!} - \frac{1}{3!} + \ldots \frac{(-1)^r}{n!}\right.$$

The total number of possible ways is $n!$ Thus the probability required is

$$\frac{1}{2!} - \frac{1}{3!} \cdots \frac{(-1)^n}{n!},$$

i.e. the first $(n-1)$ terms of $1 - e^{-1}$.

Example 7.3

Three pennies are tossed. What is the probability that they fall either all heads or all tails ?

We assume that the probability of a head with any penny is $\frac{1}{2}$ and that the result with one penny is independent of that with the others. Then there are eight possible and equiprobable cases, HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. Two of these give us all heads or all tails and hence the required probability is $\frac{1}{4}$.

Now consider this argument: there are two possibilities, either the three coins all fall alike or two of them are alike and the other different. Of these two possibilities one is of the type required and therefore the probability is $\frac{1}{2}$.

Consider also this argument: there are four possibilities, three heads, two heads and a tail, two tails and a head, three tails. Two of these four are of the type required and therefore the probability is $\frac{1}{2}$.

Finally, consider this argument: of the three coins two *must* fall alike. The other must either be the same as these two or different. Thus there are two possibilities and again the chance is $\frac{1}{2}$.

These three arguments are fallacious. They assume equiprobability among events which are not equiprobable and the application of Rule 2 is not legitimate. For example, in the first case, it is true that there are two possibilities, but they are not equal under our assumptions. The reader may care to examine why this is so and how the other two arguments break down on the same point.

Example 7.4

Peter and Paul play a game with two dice. Peter plays first by throwing the dice together. If the total number of points is a prime number other than 2 he wins outright; if it is even he throws again under the same conditions; in other cases the throw passes to Paul, who throws under the same conditions. What is the probability of Peter's winning ?

It is to be assumed that the probabilities of throwing any number 1 to 6 with either die are equal. The possible throws are 2, 3, 4 . . . 12 and the number of ways in which they can occur are :—

Total points . . .	2	3	4	5	6	7	8	9	10	11	12	Total
No. of ways . . .	1	2	3	4	5	6	5	4	3	2	1	36

Thus, according to Rule 2, the probability (1) of throwing a prime is $\frac{14}{36}$, (2) of throwing an even number is $\frac{18}{36}$, (3) of throwing neither is $\frac{4}{36}$.

These three events are mutually exclusive. Let P be the probability of Peter's winning. Now if Peter throws a prime other than 2 he wins outright, and the probability of his doing so is thus $\frac{14}{36}$; if he throws an even number he throws again, and his probability of winning in this case (according to Rule 4) is $\dfrac{18P}{36}$; if he throws neither the throw passes to Paul, whose chance is then P, so that Peter's chance of winning is $\frac{4}{36}(1-P)$.

Thus, according to Rule 3, we have

$$P = \frac{14}{36} + \frac{18P}{36} + \frac{4}{36}(1 - P),$$

giving

$$P = \frac{18}{22}.$$

7.12. It is possible to carry mathematical problems on the foregoing lines to great lengths, and a considerable amount of ingenuity has been expended in doing so. The important thing to note from the point of view of the theory of probability is that in all such cases certain probabilities are stated *a priori*, either explicitly or implicitly in some such form as " the dice are perfect " or " the selection is made at random ". One of the most formidable problems of statistics is that only in exceptional cases is there any prior certainty about the probabilities of observed events.

Probability in a Continuum

7.13. Up to this point we have considered only probabilities of finite and discrete events ; but we may also ask whether any meaning can be attached to probabilities in a continuum. For example, if a square is inscribed in a circle, what is the probability that a point taken at random in the circle is also inside the square ? If a line is divided into three segments, what is the probability that they can form a triangle ? What is the probability that $x < x_0$ where x is a positive real number less than y_0 ? And so on.

All probabilities of this kind must be considered as limits. Consider the first example, that of the square inscribed in the circle. Imagine the whole figure divided into small cells of area ε by a rectangular mesh. If we assume that the occurrence of a point in a cell is equally probable for all cells, the probability that a point falls inside both circle and square is the ratio of the number of cells in the latter to those in the former, neglecting the cells at the edges which become of diminishing importance as $\varepsilon \rightarrow 0$. In fact, the required probability can be made as near the ratio of the area of the square to that of the circle as we please by taking ε small enough. We may say that the probability *is* that ratio, which is easily seen to be $\frac{2}{\pi}$, an incommensurable number.

We should get the same limiting form of probability if we took other meshes which adequately represented areas ; but it is most important to specify the method of procession to the limit in speaking of probabilities in a continuum. Otherwise the result has no meaning. The following example will illustrate the point.

Example 7.5

Consider a straight line OA bisected at B. What is the probability that a point chosen at random on the line falls into the segment OB ?

Let us suppose in the first place that the line is divided into n equal segments of length $\frac{OA}{n}$. If we interpret the choosing of a point at random to mean the choice of one of these intervals, the probability is obviously $\frac{1}{2}$ as $n \quad \infty$, for there will be half the intervals in the segment OB.

Now let OP be drawn perpendicular to OA and equal to it in length, and imagine a star of $n + 1$ lines drawn through P, including OP and PA, so as to divide the angle $OPA\left(= \frac{\pi}{4}\right)$

into equal angles $\frac{\pi}{4n}$. These lines cut off segments on OA, and we may, if we regard equal angles as having equal probability, assign to these segments an equal probability, for they subtend equal angles at P. If we make this convention it is evident that as $n \to \infty$ the probability of a point falling into any segment on OA is proportional to the angle subtended at P. For example, the probability that a point falls in the segment OB is $\tan^{-1}\frac{1}{2}\Big/\frac{\pi}{4}$.

Now this is not the same answer that we got by assuming all small segments of OA equally probable. There is nothing paradoxical in this—the two answers are different because the two limiting processes were different. On a little reflection it will be clear that by moving the point P on the perpendicular to OA and taking a star of lines as before we can make the probability of obtaining a point in OB have any value we like. It is thus abundantly clear that the concept of probability in a continuum depends on the limiting process by which that continuum is reached from a finite subdivision of equiprobable intervals.

7.14. We have spoken above of the selection of objects "at random". In the mathematical theory of probability it is customary to define randomness in terms of probability itself. A member of a population is said to be chosen at random if it is chosen by a random method; and a random method is one which makes it equally probable that each member of the population will be chosen. Randomness is extremely important in the theory of sampling and we shall consider it at some length in the next chapter. At this point it is sufficient to note that when we speak of random choice we really mean a method of selection which gives to certain propositions an equal probability and hence allows us to apply the calculus of probability *a priori*. The justification for this is, in the ultimate analysis, empirical. It is found in practice that there exist selective processes which educe members of a population in such a way that the constituent events may be regarded as equiprobable; and the theory of sampling is largely concerned with samples generated by such processes.

It may be noted that, for continuous probabilities, randomness is dependent on the process to the limit just as probability itself is.

The Approach of von Mises

7.15. Suppose now we have a population of objects, each of which bears one of a number of characteristics. To simplify the exposition we will suppose that there are two characteristics denoted by 0 and 1. Suppose we draw members from this population and replace each member after drawing. Then the process of continued selection will generate a series such, for example, as

$$K = 011001001110101111100100 \ldots \quad (7.9)$$

Von Mises (1936) takes as the foundation of his theory of probability an infinite sequence of this kind, the Irregular Kollektiv, obeying the following laws :—

(*a*) The proportions of 0's in the first n terms tends to a limit as $n \to \infty$. This limit is called the probability of the zero in the Kollektiv.

(*b*) If a subsequence is picked out of the Kollektiv by some method which is independent of the Kollektiv itself (e.g. every third member, every member whose ordinal is

a square, every member following a zero, etc.), the limit of zeros also tends to p for $n \to \infty$; and this for every such subsequence.

The Irregular Kollektiv might, in fact, be described as the infinite random series. It has no systematic qualities; for if, for example, the series consisted of repetitions of 0110, thus

$$K = 011001100110 \ldots \qquad (7.10)$$

the subsequence consisting of every $(4r + 3)$th would consist entirely of unities and the condition (*b*) would be violated.

7.16. It is not difficult to show that probability defined in this way obeys the four rules enunciated earlier in this chapter. Some authorities have, however, found difficulty in accepting the basic concept of the Irregular Kollektiv and attributing any meaning to its existence. It has even been claimed that the idea is self-contradictory, though this von Mises strongly contests.

However this may be, the von Mises approach represents, in my own opinion, the nearest to a satisfactory basis of the frequency theory of probability that has been given. The mathematics of the subject are much the same in any of the frequency theories once the fundamental rules have been established, but when it comes to relating theory to experience the von Mises method has decided advantages. For a discussion of this subject, reference may be made to the works listed at the end of this chapter; in particular I have given (1941) the outline of a theory which in my view eliminates the difficulties associated with the Irregular Kollektiv.

Probability and Statistical Distribution

7.17. We now proceed to consider the relationship between the theory of probability and that of statistical distributions. Suppose we have a statistical population, finite and discontinuous, distributed according to a variate x. If we take a member at random from this population the probability that it bears an assigned variate-value x_0 is the frequency function $f(x_0)$, for this is the proportion of members bearing that value. Further, the probability that it bears a value less than or equal to x_0 is the distribution function $F(x_0)$, as follows at once from Rule 3 and the definition of the distribution function.

This is the essential link between probabilities and distributions. The distribution function gives the probability that a member of the population chosen at random will bear a specified value of the variate or less. We must, however, consider whether this statement can still be regarded as true for populations which are infinite or continuous.

Suppose in the first instance that the population is infinite and discontinuous. In such a case we cannot select a member at random, but we may, in the manner of **7.13**, imagine a selection from a finite population which tends to the infinite form under consideration. In this finite population the proportion of members with values less than or equal to some x_0 will be $F(x_0)$ and thus, with due regard to the nature of the limiting process, we may still say that in the infinite population the probability of a value less than or equal to x_0 is $F(x_0)$.

Similarly for a continuous distribution. In Chapter 1 we considered the continuous form as a limiting expression of

$$\Delta F = f(x)\, \Delta x.$$

If a member is chosen at random from this population in such a way that equal ranges Δx are equally probable, the probability that it falls in the range Δx is $f(x)\, \Delta x$. In the limit we may say that the probability of obtaining a value less than or equal to x_0 in taking a member

at random from a continuous population is $\int_{-\infty}^{x_0} dF = F(x_0)$. It must, however, be remembered that the nature of the process to the limit should be specified.

Hereafter, in speaking of selecting a member at random from a population $dF = f(x)\,dx$ we shall assume that what is meant is a selection random in the limit for intervals dx, i.e. such that intervals dx are equally probable.

The Concept of Random Variable

7.18. The idea of a variable x which can appear with varying degrees of probability $dF = f(x)\,dx$ has been elevated by mathematicians into a distinct concept, that of a *random variable*. In ordinary analysis no such idea appears. We write " a variable x " meaning that we are considering propositions about numbers which may be any of a certain range; there is no thought that one of these values is to be considered more frequently than others or that it will occur more frequently in practice. The random variable, on the other hand, is to be regarded as defined by a distribution function. It may take any values in a given range, but the values are distinguished by an associated function.

7.19. Let us consider what is meant by the addition of random variables. In ordinary analysis, given two variables x and y, we may define a third variable

$$z = x + y,$$

which merely means that when $x = x_0$ and $y = y_0$, z will be $x_0 + y_0$. If x and y are random variables, can we attach any useful meaning to z ?

If the joint distribution function of x and y is F_{12}, we have that the frequency of $x \leqslant x_0$ and $y \leqslant y_0$ is $F_{12}(x_0, y_0)$. Consider some value z_0. We may then determine from F_{12} the frequency such that $x + y \leqslant z_0$ which will, in fact, be the integral

$$\int\int dF_{12}(x, y)$$

taken over the region for which $x + y \leqslant z_0$.

This integral defines a function of z_0 which is in fact a distribution function, for it is zero at $-\infty$, non-decreasing, and unity at $+\infty$. We may then define this as the distribution function of the random variable z and say that z is the sum of the random variables x and y.

7.20. More generally, suppose we have n random variables distributed in the multivariate form $dF(x_1 \ldots x_n)$. We may then define a random variable z by a functional equation

$$z = z(x_1 \ldots x_n). \qquad (7.11)$$

The distribution function of z_0 is the integral of $dF(x_1 \ldots x_n)$ over all values of $x_1 \ldots x_n$ such that $z_0 \geqslant z(x_1 \ldots x_n)$. We may regard the equation (7.11) as defining a new random variable z with this as its distribution function.

Sampling Distributions

7.21. We have noted that if a member of a population is chosen at random, the probability that it will bear a variate-value not greater than x is the distribution function $F(x)$. Similarly, if we choose a member from a multivariate population, the probability that it will bear a value of the first variate not greater than x_1, of the second not greater than x_2, . . . of the nth not greater than x_n, is the multivariate distribution function

$F(x_1, x_2, \ldots x_n)$. Further, if the variates are independent, as defined in **1.33**, the rth variate being distributed as $dF_r(x_r)$, this probability is equal to

$$F_1(x_1)\, F_2(x_2) \;\ldots\; F_n(x_n).$$

Now suppose that we have a selective process, which we will call sampling, applied to a univariate population in such a way that it abstracts a group of n members. If this process is repeated it will generate a multivariate distribution, each sample exhibiting n values $x_1 \ldots x_n$. The nature of this multivariate distribution depends on the sampling process as well as the population. If the distribution is $G(x_1 \ldots x_n)$, then this function represents the probability that a random sample will result in n values, the first not greater than x_1, the second not greater than x_2, and so on.

There is one type of sampling process of outstanding importance in statistical theory, namely that in which the distribution $G(x_1 \ldots x_n)$ is the product of factors $G_1(x_1)$, $G_2(x_2) \ldots G_n(x_n)$. In such a case the sampling is said to be simple. The distributions of the values $x_1 \ldots x_n$ are independent one of another, and we may thus say that the *selection* of any member is independent of that of any other. Moreover, if the sampling is random, every $G_r(x)$ will be equal to $F(x)$, the distribution function of the population. Thus in this case we have, for the distribution of the variate-values in samples of n obtained by a simple random method,

$$\begin{aligned} dF(x_1 \ldots x_n) &= dF(x_1)\, dF(x_2) \ldots dF(x_n) \\ &= f(x_1)\, f(x_2) \ldots f(x_n)\, dx_1\, dx_2 \ldots dx_n, \qquad (7.12) \end{aligned}$$

and $F(x_1)\, F(x_2) \ldots F(x_n)$ is the probability that in such a sample the first value will not exceed x_1, and so on. Moreover, since the x's appear symmetrically in (7.12) their order is not material. The equation gives the probability that *one* member of the sample will not exceed x_1, another x_2, and so on.

7.22. Suppose now we have a sample of n members of the population with variate-values $x_1 \ldots x_n$. We may construct from these values some function, say

$$z = z(x_1 \ldots x_n), \qquad (7.13)$$

which might, for example, be the mean or the variance. We may then ask: on certain hypotheses as to the way in which the sample was derived, what is the probability that z is not greater than some assigned value z_0? In terms of frequency, if all possible samples $x_1 \ldots x_n$ were drawn and z computed for each of them, what proportion would fail to exceed some value z_0?

As an illustration, suppose we draw a sample of two from the normal population

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\, e^{-\frac{x^2}{2\sigma^2}}\, dx.$$

Let the sampling be simple and random. Then in virtue of (7.12) the probability of values in the ranges centred at x_1 and x_2 is

$$dP = \frac{1}{\sigma^2 . 2\pi} \exp - \frac{1}{2\sigma^2}(x_1^2 + x_2^2)\, dx_1\, dx_2. \qquad (7.14)$$

Consider now the quantity

$$z =$$

2

What is the probability that z shall be not greater than some assigned z_0? It is seen to be the integral of dP in equation (7.14) over the region such that $\frac{1}{2}(x_1 + x_2) \leqslant z_0$, i.e.

$$P(z \leqslant z_0) = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{x_0}\int_{-\infty}^{2z_0-x_1} \exp\left\{-\frac{1}{2\sigma^2}(x_1^2 + x_2^2)\right\}dx_1\,dx_2$$

Write

$$z = \tfrac{1}{2}(x_1 + x_2)$$
$$y = \tfrac{1}{2}(x_1 - x_2).$$

The integral becomes

$$\frac{1}{\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{z} \exp\left\{-\frac{1}{\sigma^2}(z^2 + y^2)\right\}dz\,dy$$
$$= \frac{1}{\pi\sigma^2}\int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{\sigma^2}\right)dy \int_{-\infty}^{z} \exp\left(-\frac{z^2}{\sigma^2}\right)dz$$
$$= \frac{1}{\sigma\sqrt{\pi}}\int_{-\infty}^{z} \exp\left(-\frac{z^2}{\sigma^2}\right)dz. \quad . \quad . \qquad (7.15)$$

Thus

$$P(z_0 - \tfrac{1}{2}dz_0 \leqslant z \leqslant z_0 + \tfrac{1}{2}dz_0) = \frac{1}{\sigma\sqrt{\pi}}\, e^{-\frac{z_0^2}{\sigma^2}}\, dz_0, \qquad (7.16)$$

a result which, remembering the relation between probability and the distribution function, we may express by saying that z is distributed normally with variance $\frac{1}{2}\sigma^2$. The distribution function of the statistic z is given by (7.15) and its frequency function by (7.16).

7.23. In the more general case of a statistic $z = z(x_1 \ldots x_n)$ we see that the probability of $z \leqslant z_0$ is obtained by integrating the joint distribution of $x_1 \ldots x_n$ over the domain of x's such that $z_0 \geqslant z(x_1 \ldots x_n)$. This gives us the distribution function of the random variable z defined in terms of the random variables $x_1 \ldots x_n$ by the equation $z = z(x_1 \ldots x_n)$. We shall develop this subject systematically in Chapter 10.

When the x-values are chosen by a simple random process the distribution of z is called a simple random sampling distribution, or more shortly a sampling distribution. Unless otherwise specified the words "sampling distribution" are always to be taken to refer to sampling under simple random conditions.

Bayes' Theorem

7.24. We now revert to the theory of probability. Suppose that $q_1 \ldots q_n$ are alternative propositions and let H be the information available, p some additional information. Then by Rule 4

$$P(q_r p \mid H) = P(p \mid H)\, P(q_r \mid pH)$$
$$= P(q_r \mid H)\, P(p \mid q_r H)$$

whence

$$\frac{P(q_r \mid pH)}{P(q_r \mid H)} = \frac{P(p \mid q_r H)}{P(p \mid H)}.$$

Thus

$$P(q_r \mid pH) = \frac{P(q_r \mid H)\, P(p \mid q_r H)}{P(p \mid H)}. \quad . \quad . \quad . \quad . \qquad (7.17)$$

Since the truth of one of the q's is certain we have, summing for all q's,

$$1 = \sum_{r} \frac{P(q_r \mid H)\, P(p \mid q_r H)}{P(p \mid H)}, \quad . \quad . \quad . \qquad (7.18)$$

whence, from (7.17),

$$P(q_r \mid pH) = \frac{P(q_r \mid H).P(p \mid q_rH)}{\Sigma P(q_r \mid H)\,P(p \mid q_rH)} \quad . \quad . \quad . \quad . \quad . \quad (7.19)$$

or, for variations in q_r,

$$P(q_r \mid pH) \propto P(q_r \mid H)\,P(p \mid q_rH). \quad . \quad . \quad . \quad . \quad (7.20)$$

This is Bayes' Theorem. It states that the probability of q_r on data p and H is proportional to the product of that of q_r on H and p on q_r and H.

The principal application of the theorem lies in reasoning from observed events to the hypothesis which may explain them. The theory of this subject is accordingly known as that of "inverse" probability. Suppose, in fact, that an event can be explained on the mutually exclusive hypotheses $q_1 \ldots q_n$ and let H be the data known before the event happens, so that H is the basis on which we first judge the relative probabilities of the q's. Now suppose the event to happen. Then Bayes' theorem states that the probability of q_r *after it has happened* (i.e. on data H and p) varies as the probability *before it happened* multiplied by the probability that it happens on data q_r and H. The probability $P(q_r \mid pH)$ is therefore called the *posterior* probability, $P(q_r \mid H)$ the *prior* probability, and $P(p \mid q_rH)$ will be called the *likelihood*.

In this book the word "likelihood" will be used solely in this special sense.

7.25. The practical use of Bayes' theorem depends on a knowledge of the prior probabilities. When they are known we can calculate and compare the posterior probabilities of the hypotheses, and if we have to choose one in preference to others we choose the one with the greatest posterior probability. But we are rarely, if ever, given the prior probabilities. And this brings us to what is perhaps the most contentious point in the modern theory of probability.

Bayes stated (though he appears to have felt more hesitation than most of his followers) that if there was no known reason for supposing that the prior probabilities were different, they were to be assumed equal. This is Bayes' *postulate*, which is to be distinguished from the theorem of (7.19). It immediately resolves the difficulty of applying the theorem, and before discussing the postulate and describing other approaches to the matter, it may be useful to give two examples of the use of the postulate in practical problems.

Example 7.6

An urn contains four balls, which are known to be either (a) all white, or (b) two white and two black. A ball is drawn at random and found to be white. What is the probability that all the balls are white?

We have here two hypotheses, q_1 and q_2. On q_1 the probability of getting a white ball is 1, on q_2 it is $\frac{1}{2}$. From (7.19) we have

$$P(q_1 \mid pH) = \frac{P(q_1 \mid H)}{P(q_1 \mid H) + \frac{1}{2}P(q_2 \mid H)}$$

$$P(q_2 \mid pH) = \frac{\frac{1}{2}P(q_2 \mid H)}{P(q_1 \mid H) + \frac{1}{2}P(q_2 \mid H)}.$$

Now, in accordance with Bayes' postulate we assume

$$P(q_1 \mid H) = P(q_2 \mid H) = \tfrac{1}{2}$$

and find

$$P(q_1 \mid pH) = \tfrac{2}{3}$$
$$P(q_2 \mid pH) = \tfrac{1}{3}.$$

We are thus led to prefer the hypothesis q_1 that all the balls are white, since this has the greater posterior probability.

Example 7.7

From an urn full of balls of unknown colour a ball is drawn at random and replaced. The process is continued m times and a black ball is drawn each time. What is the probability that if a further ball is drawn it will be black ?

The question as framed does not admit of a definite answer, for, there being an infinite number of possible colours and combinations of colours, we do not know what are the hypotheses which are to be compared. Let us suppose that the balls are either black or white, and thus consider the hypotheses (1) that all are black, (2) that all but one are black, (3) that all but two are black, and so on. The problem still lacks precision, for the number of balls is not specified. Suppose there are N balls. We shall later let N tend to infinity to get the limiting case.

Consider the hypothesis that there are R black balls and $N-R$ white ones. The probability of choosing a black ball is $\frac{R}{N}$ and that of doing so m times in succession, in virtue of Rule 4, is $\left(\frac{R}{N}\right)^m$. If the q's have equal prior probabilities we have, from (7.19),

$$P(q_r \mid pH) = \frac{\left(\frac{R}{N}\right)^m}{\sum_{R=0}^{N}\left(\frac{R}{N}\right)^m}.$$

Now the probability of getting a further black ball on hypothesis q_r is $\frac{R}{N}$. Since the hypotheses q are mutually exclusive, the probability of getting a further black ball is, in virtue of Rules 3 and 4,

$$\sum_{R=0}^{N}\frac{R}{N}P(q_r \mid pH) = \frac{\sum_{R=0}^{N}\left(\frac{R}{N}\right)^{m+1}}{\sum_{R=0}^{N}\left(\frac{R}{N}\right)^m}$$

This is the answer to the limited form of the question. As $N \to \infty$ this tends to the quotient of definite integrals

$$\frac{\int_0^1 x^{m+1}\,dx}{\int_0^1 x^m\,dx} = \frac{m+1}{m+2}.$$

This is a particular case of the so-called Succession Rule of Laplace. Enthusiasts have applied it indiscriminately in some such unconditioned form as the statement that if an event is observed to happen m times in succession the chances are $m+1$ to 1 that it will happen again. This is clearly unjustified.

7.26. The principal difficulties arising out of Bayes' postulate appear from the standpoint of the frequency theory of probability. If we adopt the axiomatic approach, in which

probability is a measure of attitudes of mind, it is reasonable to take prior probabilities to be equal when nothing is known to the contrary, for the mind holds them in equal doubt. The frequency theory, however, would require the states of events corresponding to the various q's to be distributed with equal frequency in some population from which the actual q has emanated, if Bayes' postulate is to be applied. This has appeared to some statisticians, though not to all, to be asking too much of the universe. The postulate is one of the crucial points in the theory of probability. Adherents of the axiomatic school accept it. Many of those of the frequency school explicitly reject it.

There is still so much disagreement on this subject that one cannot put forward any set of viewpoints as orthodox. One thing, however, is clear—anyone who rejects Bayes' postulate must put something in its place. The problem which Bayes attempted to solve is supremely important in scientific inference and it scarcely seems possible to have any scientific thought at all without some solution, however intuitive and however empirical, to the problem. We are constantly compelled to assess the degree of credence to be accorded to hypotheses on given data ; the struggle for existence, in Thiele's phrase, compels us to consult the oracles.

The Principle of Maximum Likelihood

7.27. The school of statisticians which rejects Bayes' postulate has substituted for it an apparently different principle based on the use of likelihood. Reverting to equation (7.19) we see that for any q_r and H

$$P(q_r \mid pH) \propto P(q_r \mid H)\, L(p \mid q_r H), \qquad (7.21)$$

where we now write $L(p \mid q_r H)$ for the likelihood function. The Principle of Maximum Likelihood states that when confronted with a choice of hypotheses q we are to select that one (if it exists) which maximises $L(p \mid q_r H)$. In other words, we are to choose the hypothesis which gives the greatest probability to the observed event.

It is to be particularly noted that this is not the same thing as choosing the hypothesis with the greatest probability. In fact, some adherents of the frequency theory of probability deny any meaning to the expression " probability of a hypothesis ", and the principle of maximum likelihood was introduced largely to replace the notion of " inverse " probability which leads to the use of such a phrase.

7.28. Suppose (as is nearly always the case in statistical work) that the hypotheses with which we are concerned assert something about the numerical value of a parameter θ. In such a case we shall speak of a statistical hypothesis. For instance, the hypotheses might be $q_1 \equiv \theta < 0$, $q_2 \equiv \theta \geqslant 0$, in which case there are two alternatives. Or we might have $q_1 \equiv \theta = 1$, $q_2 \equiv \theta = 2$, and so on, in which case there is a denumerable infinity of hypotheses.

If now θ can have only discontinuous values, we may, confronted with an observed event p, require to estimate θ, or to ask what is the " best " value of θ to take on the evidence p. The method of Bayes would state that the " best " value was the most probable value. In (7.21) we should seek for that q_r which made $P(q_r \mid pH)$ a maximum. If we know nothing of the prior probabilities $P(q_r \mid H)$ we should, in accordance with Bayes' postulate, assume all such probabilities equal. We then merely have to find that q_r which maximises $L(p \mid q_r H)$. In other words, the postulate of Bayes and the principle of maximum likelihood result in the same answer and are equivalent.

7.29. This position apparently does not hold if the permissible values of θ are continuous. We must now replace such expressions as $P(q_r \mid H)$ by $P(\theta_0 - \frac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 - \frac{1}{2}d\theta_0 \mid H)$ and in place of (7.21) we get

$$P(\theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0 \mid pH) \propto P(\theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0 \mid H) \times L(p \mid \theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0, H). \quad (7.22)$$

If we now require the " best " value of θ, we should, in accordance with Bayes' postulate, take the prior probability to be a constant and once again we should have to maximise L for variations of θ.

We might, however, have chosen to represent our hypotheses, not by θ, but by some variate ϕ functionally related to θ, e.g. the standard deviation instead of the variance. In this case we should have reached equation (7.22) with ϕ written everywhere instead of θ; we should have taken the prior probability as constant; and we should have arrived at the conclusion that we should maximise L for variations of ϕ.

But are we being consistent in so doing? If we assume that the elementary intervals of θ are equiprobable we cannot assume the same of ϕ, and thus the use of Bayes' postulate appears to involve self-contradiction. The principle of maximum likelihood is free from this difficulty, for if $L(\theta)$ is to be maximised for variations of θ it will, at the same time, be maximised for variations of ϕ, since

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \phi}\frac{\partial \phi}{\partial \theta}$$

and the two sides of this equation vanish together.

7.30. This is one of the grounds on which adherents of the frequency school have rejected Bayes' postulate in favour of the principle of maximum likelihood; but in my view the matter has been misunderstood. It would seem that Bayes' postulate and the principle give the same answer in the continuous case as well as in the discontinuous case when proper regard is had to the limiting processes involved. We saw in **7.13** that in speaking of probability in a continuum it was essential to specify the nature of the process to the limit. If we regard θ (from the frequency viewpoint) as having emanated from a population by a process random in the limit for intervals $d\theta$, then Bayes' postulate applied to this process will clearly give a different answer from that obtained by supposing that θ emanated by a process random in the limit for $d\phi \left(= \frac{\partial \phi}{\partial \theta} d\theta\right)$. The two are different just as the probabilities in Example 7.5 are different, and for the same reason. Thus the apparent inconsistency is not an inconsistency at all, but a difficulty introduced by ignoring the limiting process in continuous populations.*

For an extended discussion of this subject reference may be made to Kendall (1940). In the present volume it need not concern us to take it farther, though considerable use will be made of the principle of maximum likelihood in Volume 2. It will there be seen that the principle has many important statistical properties. No one, in fact, denies the importance

* A further difficulty arises if θ can lie in an infinite range, for then Bayes' postulate apparently leads to the conclusion that prior probabilities in any finite range are zero and hence so are posterior probabilities. This does not arise in the likelihood method. Jeffreys overcomes it by assuming that the prior probability in such a case is inversely proportional to the parameter θ. Looking at the problem generally, we need not be surprised that the difficulty appears since the ranging of θ over an infinite range is also a limiting process. In practice we are never so ignorant *a priori* as to suppose that θ can be any value however large with the same probability, and if we consider the range as determinate but unknown, likelihood and Bayes' postulate continue to be applicable and to give the same results.

of the principle or its usefulness in certain cases ; the controversy hitherto has centred on the considerations by which the acceptance of the principle as a rule of conduct is to be justified. The reader who cannot accept Bayes' postulate and the foregoing argument that it is virtually identical with the principle has a choice of courses. He can accept the principle as a new and distinct postulate of scientific inference ; he can regard it as justified by its mathematical and statistical properties ; or he can rely on a more sophisticated approach which will be touched on in Chapter 9, namely, that the principle leads to estimates of parameters with minimum sampling variance when such exist. At this stage he may be prepared to accept it on intuitive grounds.*

7.31. Although in the remainder of the present volume Bayes' postulate and the principle of maximum likelihood will not often appear explicitly, we shall frequently use a type of argument which is, in the ultimate analysis, based on them. A certain event or series of events is observed ; on a hypothesis H the occurrence of these events is found to be highly improbable ; and therefore H is rejected in favour of some hypothesis which makes the observations more probable. To take a very simple example, we toss a penny twenty times and find that it comes down heads every time. If the penny were unbiased (hypothesis H) the odds against this event would be $2^{20} - 1$ to 1. Thus we reject H in favour of the hypothesis that it is in fact biased in favour of the heads.

It will readily be seen that this type of argument is a somewhat indefinite form of the inverse type with which we have been concerned. The chief difference lies in the fact that it is used to reject unlikely hypotheses rather than to accept the most likely, possibly a safer but certainly a less precise procedure.

The Central Limit Theorem

7.32. To conclude this chapter we prove an important theorem which gives the normal distribution a central place in the theory of probability and the theory of sampling. It has already been shown that the distribution appears as the limiting form of the binomial and the Pearson Type III distribution when expressed in standard measure. We shall prove a much more general result, due to Laplace but first proved rigorously by Liapounoff, that under certain conditions the sum of n independent random variables distributed in whatever form tends, when expressed in standard measure, to the normal form as n tends to infinity. This is the famous Central Limit Theorem.

Let us note in the first place a simple but powerful result connected with the characteristic functions of sums of independent random variables. If we have n such variables distributed as $dF_1 \ldots dF_n$ the element of frequency of their sum $z = x_1 + \ldots x_n$ is the integral of $dF_1 \ldots dF_n$ through the element of volume between z and $z + dz$. Thus the characteristic function of their sum, being the integral of e^{itz} through the range of z, is equal to

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{itz}\, dF_1 \ldots dF_n$$

$$= \int_{-\infty}^{\infty} e^{itx_1}\, dF_1 \int_{-\infty}^{\infty} e^{itx_2}\, dF_2 \ldots \int_{-\infty}^{\infty} e^{itx_n}\, dF_n$$

$$= \phi_1 \phi_2 \ldots \phi_n.$$

* An approach of a rather different kind has been developed in recent years by Neyman (1937), who bases his theory of inference only on direct probabilities. An account of this theory will be given in the second volume.

That is to say, the characteristic function of the sum of a number of independent random variables is the product of their characteristic functions. The cumulative function is accordingly the sum of their cumulative functions.

Now as to the Central Limit Theorem itself. We first of all outline the proof briefly and unrigorously to indicate its essential features, and then give a rigorous proof. Suppose we have distributions $F_1 \ldots F_n$, all with finite second moments and with characteristic functions $\phi_1 \ldots \phi_n$. We have for any F_r

$$\phi_r(t) = 1 + \mu_1'\frac{(it)}{1!} + \mu_2'\frac{(it)^2}{2!} + R_r$$

when R is a remainder term. Similarly we have

$$\psi_r(t) = \mu_1'\frac{(it)}{1!} + \mu_2\frac{(it)^2}{2!} + R_r.$$

Hence the cumulative function of the sum of the independent variates will be

$$\Psi(t) = (it)\Sigma\mu_1' + \frac{(it)^2}{2!}\Sigma\mu_2 + \Sigma R.$$

We can without loss of generality take the mean of the sum as origin, so that $\Sigma\mu_1' = 0$, and now transforming to standard measure by the transformation $\xi = \frac{x}{\sqrt{\Sigma\mu_2}}$ we find

$$\Psi(t) = \frac{-t^2}{2!} + O\left(\frac{\Sigma R}{(\Sigma\mu_2)^{\frac{3}{2}}}\right).$$

Since $\Sigma\mu_2$ is of order n the remainder term will be of order $\frac{n}{n^{\frac{3}{2}}}$, i.e. of order $n^{-\frac{1}{2}}$; and thus tends to zero. We shall then have

$$\lim \Psi(t) = -\frac{t^2}{2}$$

$$\lim \Phi(t) = e^{-\frac{t^2}{2}}$$

and hence in virtue of the converse of the First Limit Theorem (**4.12**) the distribution of the sum of the random variables tends to normality.

7.33. The rigorous enunciation of the theorem and its proof are as follows :—

If n independent random variables are distributed in the forms $F_1 \ldots F_n$ with finite variances $\mu_{2,1} \ldots \mu_{2,n}$ and $M_n = \sum_{j=1}^{n}\mu_{2,j}$, then the sum of the variables divided by $\sqrt{M_n}$ tends to the normal form, provided that for any $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{1}{M_n}\sum_{j=1}^{n}\int_{|x|>\varepsilon\sqrt{M_n}} x^2\, dF_j = 0. \qquad (7.23)$$

The implications of this condition, which is a modification by Cramér of one due to Lindeberg, are not very obvious, but it involves that

$$\sqrt{M_n} \to \infty \quad \text{and} \quad \frac{\mu_{2,j}}{M_n} \to 0, \qquad . \quad . \quad . \quad . \quad . \quad (7.24)$$

in other words, that the total variance tends to infinity but that the proportional contribution of each constituent tends to zero. To see that (7.24) follows from (7.23) we note that if M_n does not tend to infinity it must, being an increasing function, tend to a constant. It would

follow from (7.23) that the sum of the integrals, each of which is positive and not small for every ε, would tend to zero, which is impossible. Further, if $\frac{\mu_{2,j}}{M_n}$ did not tend to zero, then at least one of the terms in (7.23) would not do so, and thus the sum would not do so.

We have

$$\phi_j\left(\frac{t}{\sqrt{M_n}}\right) = \int_{-\infty}^{\infty} e^{\frac{itx}{\sqrt{M_n}}}\,dF_j$$

$$= \int_{|x|>\varepsilon\sqrt{M_n}} + \int_{|x|\leqslant\varepsilon\sqrt{M_n}} e^{\frac{itx}{\sqrt{M_n}}}\,dF_j.$$

Expanding the exponential with a Maclaurin remainder we have

$$\phi_j\left(\frac{t}{\sqrt{M_n}}\right) = \int_{|x|>\varepsilon\sqrt{M_n}}\left(1+\frac{itx}{\sqrt{M_n}}+\theta\frac{(it)^2x^2}{2M_n}\right)dF_j$$
$$+\int_{|x|\leqslant\varepsilon\sqrt{M_n}}\left(1+\frac{itx}{\sqrt{M_n}}+\frac{(it)^2x^2}{2M_n}+\theta'\frac{(it)^3x^3}{6M_n^{\frac{3}{2}}}\right)dF_j \qquad 0\leqslant|\theta|,|\theta'|\leqslant 1.$$

We may without loss of generality suppose the mean to be zero and hence we find

$$\phi_j\left(\frac{t}{\sqrt{M_n}}\right) = 1-\frac{t^2}{2}\frac{\mu_{2,j}}{M_n}+\frac{\theta}{2M_n}\int_{|x|>\varepsilon\sqrt{M_n}} t^2x^2\,dF_j$$
$$+\frac{\theta'}{6M_n^{\frac{3}{2}}}\int_{|x|\leqslant\varepsilon\sqrt{M_n}} t^3|x|^3\,dF_j \qquad 0\leqslant|\theta|,|\theta'|\leqslant 1.$$

Thus for some $T>1$ we have, for $|t|<T$, remembering that

$$\int_{|x|\leqslant\varepsilon\sqrt{M_n}} dF \leqslant \varepsilon\sqrt{M_n}\int_{|x|\leqslant\varepsilon\sqrt{M_n}} x^2\,dF$$

$$\phi_j\left(\frac{t}{\sqrt{M_n}}\right) = 1-\frac{t^2}{2}\cdot\frac{\mu_{2,j}}{M_n}+\theta''\frac{T^3}{M_n}\left(\varepsilon\mu_{2,j}+\int_{|x|>\varepsilon\sqrt{M_n}} x^2dF_j\right) \qquad 0\leqslant|\theta''|\leqslant 1.$$

Hence, in virtue of (7.24) the coefficient of θ'' is as small as we please and thus $\phi_j\left(\frac{t}{\sqrt{M_n}}\right)$ tends to unity as $n\to\infty$ uniformly for $|t|<T$. Thus we have

$$\psi_j\left(\frac{t}{\sqrt{M_n}}\right) = (1+\eta)\left\{\phi_j\left(\frac{t}{\sqrt{M_n}}\right)-1\right\}$$

for sufficiently large n and $|\eta|<\varepsilon$. Thus for $\varepsilon<\frac{1}{2}$,

$$\psi_j\left(\frac{t}{\sqrt{M_n}}\right) = \frac{-t^2}{2}\frac{\mu_{2,j}}{M_n}+\frac{2\theta''T^3}{M_n}\left(\varepsilon\mu_{2,j}+\int_{|x|>\varepsilon\sqrt{M_n}} x^2dF_j\right).$$

Summing for j we have, in virtue of (7.23),

$$\Psi\left(\frac{t}{\sqrt{M_n}}\right) = \frac{-t^2}{2}+2\theta''T^3(\varepsilon+\text{vanishing quantity})$$

and thus for $|t|<T$

$$\lim\Psi\left(\frac{t}{\sqrt{M_n}}\right) = \frac{-t^2}{2},$$

the convergence being uniform in any finite t-interval. The theorem follows from the converse of the First Limit Theorem.

7.34. The following comments will amplify the above proof.

(*a*) The Lindeberg condition (7.23) is necessary as well as sufficient. A proof is given by Cramér (1937).

(*b*) The condition may be put in other forms, for which see Cramér (1937), Uspensky (1937) and the original memoir by Liapounoff (1901).

(*c*) The sum of random variables whose distributions have not a finite second moment may not tend to normality. It will be seen in Chapter 9 that the mean of n variables each of which is distributed in the form

$$dF = \frac{k\,dx}{1+x^2} \qquad -\infty \leqslant x \leqslant \infty$$

is also distributed in that form, however large n may be.

(*d*) Liapounoff has also given some remarkable results showing how close the limiting form is to the sum of n variables. In fact, if F_n is the distribution function of the sum, and F that of the normal form

$$F_n - F \mid \leqslant c\rho_{3n} \cdot \frac{\log n}{\sqrt{n}},$$

where c is a constant, ρ_{3n} is a function of the third moments of the constituent distributions.

NOTES AND REFERENCES

The logic of the theory of probability will be found dealt with in the books by Keynes (1921), F. P. Ramsey (1931) and Johnson (1921). All these take the axiomatic approach from probability as an undefined idea. The frequency approach has been discussed from the more logical angle by Venn (1888), whose book, though out of print and to some extent out of date, is still worth reading.

The mathematical theory of probability has been treated by Lévy (1925), Jeffreys (1939) and Uspensky (1937), all three books excellent of their kind. Von Mises' approach is described in his book (1936) and an axiomatisation in a paper by Dörge (1934). See also Kendall (1941).

For inverse probability and likelihood see the review by Kendall (1940). There are scores of papers, mostly controversial in character, on this subject, but a beginning of a systematic reading may be made with the papers by Fisher (1921, 1930), Neyman (1937), and the book by Jeffreys (1939).

For the central limit theorem see Cramér (1937), and for an extension to the case when the variables are dependent, Bernstein (1927).

Bayes, T. (1763), "An essay towards solving a problem in the doctrine of chances," *Phil. Trans.*, **53**, 370.

Bernstein, S. (1927), "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Math. Ann.*, **97**, 1.

Borel, E. (editor) (1925 and later years), *Traité du Calcul des Probabilités et de ses Applications*, Gauthier-Villars, Paris.

Cramér, H. (1937), *Random Variables and Probability Distributions*. Cambridge University Press.

Dörge, K. (1934), "Eine Axiomatisierung der von Misesschen Wahrscheinlichkeitstheorie," *Jber. dtsch. Mat. Ver.*, **43**, 39.

Fisher, R. A. (1921), "On the mathematical foundations of theoretical statistics," *Phil. Trans.*, A, **222**, 309.

—— (1930), "Inverse Probability," *Proc. Camb. Phil. Soc.*, **26**, 528.

Jeffreys, H. (1939), *The Theory of Probability*, Cambridge University Press.

Johnson, W. E. (1921–24), *Logic* (3 volumes), Cambridge University Press.

Kendall, M. G. (1940), "On the method of maximum likelihood," *J. Roy. Statist. Soc.*, **103**, 388.

—— (1941), "A theory of randomness," *Biometrika*, **32**, 1.

Keynes, J. M. (1921), *A Treatise on Probability*, Macmillan.

Kolmogoroff, A. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin.

Lévy, P. (1925), *Calcul des Probabilités*, Gauthier-Villars, Paris.

Liapounoff, A. (1901), "Nouvelle forme du théorème sur la limite de probabilité," *Mém. Acad. Sci. St. Pétersbourg*, **12**, No. 5.

von Mises, R. (1936), *Wahrscheinlichkeit, Statistik und Wahrheit*, Springer, Berlin. (English translation, 1939, as *Probability, Statistics and Truth*, W. Hodge.)

Neyman, J. (1937), "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans.*, A, **236**, 333.

Ramsey, F. P. (1931), *The Foundations of Mathematics*, Kegan Paul.

Uspensky, J. V. (1937), *Introduction to Mathematical Probability*, McGraw-Hill, New York and London.

Venn, J. A. (1888), *The Logic of Chance*, Macmillan.

EXERCISES

7.1. If each of an aggregate of N objects can possess or not possess any of n characteristics A, B, . . . K; and if $(ab\ .\ .\ .\ f)$ is the number of objects possessing A, B . . . F, show that the number of objects possessing at least one of A, B, . . . K is

$$\Sigma(a) - \Sigma(ab) + \Sigma(abc) \ . \ . \ . + (-1)^{n-1}\Sigma(ab\ .\ .\ .\ k).$$

In each of a packet of cigarettes there is one of a set of cards numbered from 1 to n. If a number N of packets is bought at random, the population of packets is large and the numbers are equally frequent, show that the probability of getting a complete set of cards is

$$1 - \binom{n}{1}\left(\frac{n-1}{n}\right)^N + \binom{n}{2}\left(\frac{n-2}{n}\right)^N \ . \ . \ . + (-1)^{n-1}\binom{n}{n-1}\left(\frac{1}{n}\right)^N$$

7.2. Three points are taken at random on a circle. Show that the probability that they lie on the same semi-circle is $\frac{3}{4}$. (Assume that in the limit elementary intervals of arc are equiprobable.)

Explain the fallacy in the following argument: One pair of the points *must* lie on a semicircle terminating at one of them. The probability that the third point lies on this semicircle is $\frac{1}{2}$, which is therefore the required answer.

7.3. A simple random sample of n values, x_1 . . . x_n, is drawn from the normal population

$$dF = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx.$$

Show that the value of m which maximises the likelihood of this event is

$$m = \frac{1}{n}\Sigma(x),$$

which is therefore the "best" estimate of the mean of the population.

7.4. Show that if p is the probability of a zero in the Irregular Kollektiv the probability u_n that there will be r consecutive zeros in a set of n members chosen at random obeys the recurrence relation

$$u_{n+1} = u_n + (1 - u_{n-r})p^r(1 - p)$$

and hence that

$$u_n = 1 - v_{n,r} + p^r v_{n-r,r}$$

where

$$v_{n,r} = \sum_{j=0}^{\left[\frac{n}{r+1}\right]} (-1)^j \binom{j}{n - jr} \{p^r(1 - p)\}^j.$$

CHAPTER 8

RANDOM SAMPLING

The Sampling Problem

8.1. In the previous chapter we have referred incidentally to the sampling problem, which can be stated quite simply : given a sample from a population, to determine from it the properties of that population. We noted that only in exceptional cases is it possible to make assertions about the population with complete certainty, and that consequently it is necessary to fall back on statements of a less categorical kind expressible in terms of probability.

8.2. In order to be able to apply the theory of probability to this problem it is necessary that the sampling should be random. In actual practice we often meet with samples which are not random, having been chosen purposively for some reason or other. In such circumstances it is not, as a rule, possible even to make precise statements in probability ; and where a decision has to be taken one is forced to rely on subjective judgments of an unsatisfactory kind. No numerical estimate of the probabilities can be made. It is for this reason that random sampling becomes of primary importance in statistical investigations from sample to population. From this point onwards we shall deal only with random samples, and to avoid constant repetition shall leave it to be understood that where a " sample " or a " sampling distribution " is referred to, random conditions are assumed.

8.3. It is useful to begin a discussion of random sampling by considering the types of parent population from which samples can be chosen.

(*a*) In the first place, the population may be finite and existent, e.g. the population of human beings in Europe at a fixed point of time, or the population of apples on a given tree. A sampling process which extracts members one at a time from this population will evidently eventually exhaust the supply of members if continued long enough. Thus the sampling, though random, is not simple in the sense of **7.21,** for the probability of a given member being chosen varies according to what has already been abstracted.

We may, however, reduce this process to one of simple sampling by replacing the members after withdrawal. The population then remains the same at each trial. The two cases are sometimes distinguished as " sampling without replacement " and " sampling with replacement ".

Furthermore, we may also in many cases regard the sampling as simple to an adequate approximation even when there is no replacement. If the population is large compared with the size of the sample, the abstraction of relatively few members will not materially affect the constitution of the remaining population, which may thus be regarded as approximately the same for subsequent samplings.

(*b*) Sampling with replacement from a finite population may, in fact, be regarded as sampling from an infinite population, for the process will never exhaust the supply. We may, however, have to deal with a population which is infinite in rather a different sense, namely, that of a limiting form. We may, for example, wish to consider the probability of a sample from the positive integers or the real numbers from 0 to 1. The latter case presents itself in sampling from a continuous frequency-distribution which we must necessarily regard as infinite.

Thus, if we replace an observational distribution by a conceptual continuous mathematical distribution, we replace at the same time a finite population by an infinite population. The drawing of random samples from such a population is attended by the circumstances referred to in **7.13** and **7.29**, namely, that the process to the limit must be taken into account.

(*c*) Thirdly, the population may be purely hypothetical. Consider, for example, the throws of a die. We may picture the continual throwing as a sampling process drawing existent members from some non-existent population. In such cases what we are really doing is constructing by mental fiction an imaginary population round the sample.

The concept of the hypothetical population is necessitated by ideas of frequency in probability. It is not required (and indeed has been explicitly rejected by Jeffreys) in the approach which takes probability as an undefinable measurement of attitudes of doubt. But if we take probability as a relative frequency, then to speak of the probability of a sample such as that given by throwing a die or growing wheat on a plot of soil, we must consider the sample against the background of a population. There are obvious logical difficulties in regarding such a sample as a selection—it is a selection without a choice—and still greater difficulties about supposing the selection to be random ; for to do so we must try to imagine that all the other members of the population, themselves imaginary, had an equal probability of assuming the mantle of reality, and that in some way the actual event was chosen to do so. This is, to me at all events, a most baffling conception. At the same time, it has to be admitted that certain events such as dice-throwing do happen as if the constituents were chosen at random from an existent population, and it accordingly seems that the concept of the hypothetical population can be justified empirically.

Randomness in Sampling

8.4. In its colloquial use the word "random" is applied to any method of choice which lacks aim or purpose. We speak of drawing names at random out of a hat, choosing plants at random from a field of corn, selecting family budgets at random from the population, meaning thereby that the selection is completely haphazard.

Now it is found in practice that choice by a human being is not random in the stricter sense that it produces equally frequently events which we are entitled to expect to have equal prior probabilities. Some examples will make this clear.

Example 8.1

In the course of certain work at the Rothamsted Experimental Station sets of eight wheat plants were chosen for measurement. Six of these were chosen by approved methods,

TABLE 8.1

Distribution of Plants chosen haphazardly in Ranks 1 to 8.

(F. Yates, *Ann. Eugen. Lond.*, **6**, 202.)

Date.	Observation.	Numbers bearing Specified Rank.								Total.
		1	2	3	4	5	6	7	8	
May 31st	Shoot height . .	9	7	11	8	11	18	21	31	116
June 28th	Ear height . .	9	19	27	23	15	10	5	4	112

referred to below, and may be taken to be truly random. The other two were chosen haphazardly by eye. If, in any set, the eight plants were ranged in order of magnitude, the two selected by eye could have any number from one to eight; and if they, in common with the other six, were chosen at random, they should occupy these places with approximately equal frequency in a large number of sets. Table 8.1 shows what actually occurred on two different occasions (*a*) on May 31st, before the ears of wheat had formed, and (*b*) on June 28th, after the ears had formed.

The divergence of actual from expected results is quite striking. On May 31st, before the ears had formed, the observer was strongly biased towards the taller shoots; whereas in June he was biased strongly towards the central plants and avoided short and tall plants.

Thus it is seen that bias can appear even in a trained observer, and that the bias need not be consistent in over- or under-estimation in different circumstances.

Example 8.2

The following table shows the frequencies of final digits in a number of measurements made by four different observers:—

TABLE 8.2

Bias in Scale Reading. Distribution of Final Digits in Measurements by Four Observers.

(G. U. Yule, *J.R. Statist. Soc.*, **90**, 570.)

Final Digit.	Frequency of Final Digit per 1000.			
	A	B	C	D
0	158	122	251	358
1	97	98	37	49
2	125	98	80	90
3	73	90	72	63
4	76	100	55	37
5	71	112	222	211
6	90	98	71	62
7	56	99	75	70
	126	101	72	44
	129	81	65	16
TOTAL	1001	999	1000	1000

It is hard to suppose that there was any genuine difference which would lead to the appearance of certain digits at the expense of others, and we may confidently suppose that the deviations from approximate equality indicate bias on the part of the observer.

Observer A had decided preference for 0, 2, 8 and 9, avoiding the centre of the scale. Observer B is quite good, his deviations from expected values being small, though he also showed some preference for 0. Observer C was poor, rounding off one measurement in two to the whole or half unit. Observer D was obviously very bad indeed, nearly 57 per cent. of his measurements being rounded off to the whole or half unit.

The observations were all made by reading a scale, those under A being on drawings to

the nearest tenth of a millimetre, those under B, C, and D being measurements on the heads of living subjects to the nearest millimetre. We may conclude from this that different observers may exhibit different degrees of bias even under comparable circumstances, and that even those who are aware of the existence of the possibility of bias and the necessity for taking great care (as observer A was) may nevertheless fail to avoid it.

Example 8.3

An observer was placed before a machine consisting of a circular disc divided into ten equal sections in which were inscribed the digits 0 to 9. The disc rotated at high speed and every now and then a flash occurred from a nearby electric lamp of such short duration that the disc appeared at rest. The observer had to watch the disc and write down the number occurring in the division indicated by a fixed pointer.

This was a machine designed for the provision of truly random numbers (see below, **8.10**) and had been found by another observer to do so. But this particular observer produced a definite bias. The frequencies of digits in 10,000 run off by him are shown in Table 8.3.

TABLE 8.3

Distribution of Digits obtained by an Observer in using a Randomising Machine.
(Kendall and Babington Smith, *Supp. J.R. Statist. Soc.*, 6, 51.)

Digit.	0	1	2	3	4	5	6	7	8	9	TOTAL.
Frequency . .	1083	865	1053	884	1057	1007	1081	997	1025	948	10,000

If the observer was unbiased the digits should appear in approximately equal numbers ; but there is a bias in favour of all the even numbers and against the odd numbers 1, 3 and 9. The cause of this bias is obscure, for the observer did not have to estimate (as in the previous example) but merely to write down something which he saw, or thought he saw. The explanation seemed to be that he had a strong number-preference, i.e. that he actually mis-saw the numbers, or that his brain controlled his ocular impressions and censored them. We have here to deal with one of the deadliest forms of bias in psychology.

Example 8.4

Every year a number of crop reporters in England and Wales estimate the prospective yields of certain crops, forecasts being obtained at different periods of the year and final estimates when the crop is harvested. Table 8.4 shows the average estimated yield of potatoes at the various times for the years 1929–1936.

This table exhibits very clearly an effect which has shown itself in nearly all the English crop reports (and appears also in other countries), namely, the chronic pessimism of crop forecasts. In every case but one in the above table the forecasts are below the final yield. Nor do crop reporters seem able to learn by experience that they are underestimating. Nothing in this table indicates that the differences between forecast and final estimate diminished during the period concerned.

It should also be noticed that these estimates are the weighted average of a large number of independent observations. One of the commoner misunderstandings in this type of work

TABLE 8.4

Bias in Crop Forecasting. Forecasts of Yields of Potatoes in England and Wales (Tons per Acre).

(From the official agricultural statistics.)

Year.	Sept. 1st.		Oct. 1st.		Nov. 1st.		Final Estimate.
	Yield.	% Difference from Final.	Yield.	% Difference from Final.	Yield.	% Difference from Final.	
1929	5·7	− 17·4	6·2	− 10·1	6·5	− 5·8	6·9
1930	6·0	− 7·7	6·1	− 6·2	6·1	− 6·2	6·5
1931	5·5	0·0	5·3	− 3·6	5·3	− 3·6	5·5
1932	6·4	− 3·0	6·2	− 6·1	6·3	− 4·5	6·6
1933	6·4	− 4·5	6·2	− 7·5	6·4	− 4·5	6·7
1934	6·0	− 15·5	6·3	− 11·3	6·7	− 5·6	7·1
1935	5·6	− 9·7	5·7	− 8·1	6·0	− 3·2	6·2
1936	6·0	− 3·2	5·9	− 4·8	5·8	− 6·5	6·2

is based on the supposition that, though individuals may make mistakes, their errors will cancel out in the aggregate. Our present example shows this to be untrue in general. There can appear a systematic bias affecting all the individuals performing estimates.

8.5. The foregoing examples are enough to indicate that human bias is very prevalent. Trained observers may be biased even when conscious of their own imperfections ; different observers may be biased in different ways in similar circumstances ; and the same observer may be biased in different ways in different circumstances. It is abundantly clear that we must look for true randomness elsewhere than in mere lack of purpose on the part of human observers. There may be persons whose psychological processes are so finely balanced that they can deliberately select random samples, but few statisticians who have experimented in this interesting field would regard themselves as among them.

8.6. In Chapter 7 we saw that the primary function of randomness in probability was that it ensured that certain primitive events were equally probable. We may say that a method of selection is random for a population U if, when applied to U, it gives all members an equal probability of being chosen ; or, in the language of frequency, if, when continually applied to U, it educes the members approximately equally frequently.

But this is not enough. Suppose we had a population of two members A and B, and sampled with replacement. Then a method which chooses A and B alternately and produces the series $ABAB$ . . . educes each member approximately equally frequently ; but it is not what we customarily mean by a random method. What we require of a random method is that in such circumstances it should produce a series like that of von Mises (**7.15**) in which no systematic arrangement is evident. Not only single characteristics, but all possible groups of characteristics should appear equally frequently.

8.7. A further point is to be noted. We may, in drawing the sample, be interested in one particular variate exhibited by the members, and it is possible that a method may give a satisfactory random sample *so far as this variate is concerned* without doing so for other

variates. Suppose, for example, we are anxious to take a random sample from the inhabitants of a particular street. If we are concerned with a variate such as eye-colour it might be sufficient to choose a house every so often, say every tenth house, and take the inhabitants of that house as part of the sample. Such a method would not give every inhabitant an equal chance of being chosen; but if we look back to the time when the inhabitants took up residence we may imagine that the colour of their eyes did not influence their geographical distribution, and thus that if we consider the allocation of the inhabitants in some way independent of eye-colour, and then take every tenth house, we may suppose that so far as eye-colour is concerned the sample is random. But the matter would stand differently if we were sampling for income. If for instance every tenth house was a corner house and thus inhabited by a person of more than average income, our sample would no longer be random with respect to income. Looking back, as before, to the time when inhabitants took up residence, we see that they can no longer be regarded as distributed at random, for those with larger incomes will tend to be attracted towards the more expensive houses.

Thus a method which is random for one population may not be so for another; and even in the same population a method random for one variate may not be so for another. Randomness is relative.

The Technique of Random Sampling

8.8. Suppose, then, that we are given a population and a variate is specified. How are we to draw a random sample, i.e. how can we find a method which is random for that population and that variate? The answer lies partly in theory and partly in practice.

(*a*) In the first place we must require that there is no obvious connection between the method of selection and the properties under consideration. The method and the properties must be independent so far as our prior knowledge is concerned. In sampling a field of wheat for shoot height, for example, we must not use a method which could be influenced by that height, such as skimming a hoop over the field and selecting the plants round which it fell (for the hoop might tend to catch on the taller plants). Again, in sampling the inhabitants of a town by choosing names from a telephone directory we should undoubtedly tend to get the more well-to-do classes and hence, if the variate under consideration is wealth or any related characteristic such as number of children, political opinion, standard of education and so on, the sample would not be random. If we were concerned with characteristics such as height, hair colour, or blood group the sample might be random, though it is not difficult in many similar cases to think of reasons why the variate might be linked with wealth.

If this matter is viewed from the standpoint of the axiomatic theory of probability the absence of knowledge about relationship between the method and the characteristic under consideration may be sufficient to ensure randomness, for the probabilities of elementary propositions then become equal—the probabilities being measures of prior attitudes of mind.* But if the frequency viewpoint is adopted it is not enough that there should be absence of knowledge of this kind, for unknown to the observer there may be relations which will prevent the elementary propositions from being true in approximately equal proportions. The presumption is that if we make as great an effort as possible to ascertain whether any relationship exists and fail to find it, there is no relationship; and hence we can assume randomness

* At least, this is my interpretation of the position; but the writers on the axiomatic theory have not discussed randomness at any length, being content to define it in terms of probability, and I may be putting a gloss on their views which they would not accept.

with more or less confidence. But in this approach the assumption of randomness is ultimately part of the general uncertainty of the inference from sample to population.

(*b*) Secondly, we may rely on previous experience of a random method to justify its use on new occasions. This is evidently an extrapolation, and though most people would regard it as reasonable, the fact has to be realised. The axiomatic theory of probability can embrace this extrapolation within its scope, for the probabilities given by the method are assessable in terms of prior knowledge; but the frequency theory has to take the extrapolation as an additional assumption.

8.9. One of the most reliable methods of drawing random samples consists of constructing a model of the population and sampling from the model. We may, for instance, note down the characteristics of each member on a card and sample by choosing cards from the pack corresponding to the whole population. This is the method adopted in lotteries and the process is known as lottery or ticket sampling. It is moderately effective but suffers in practice from two disadvantages: the labour of constructing the card population, and the danger of bias in the drawing of cards. Example 12.1 below, for instance, shows that the ordinary processes of shuffling and dealing playing-cards may fail to be satisfactory. To be reasonably satisfied about the randomness of the shuffling entails a good deal of trouble and labour, and the same object can be attained much more simply by the use of random sampling numbers, which we now consider.

Random Sampling Numbers

8.10. The easiest way of constructing a miniature population is to attach an ordinal number to each member, mostly simply by numbering the members from 1 onwards. The set of ordinals so obtained is the miniature population and the problem of drawing a random sample reduces to finding a series of random numbers. The advantages of this method are obvious: no physical model population has to be constructed; the numbering can be carried out in any convenient manner; and the series of random numbers can be applied to any enumerable population so that any series of random numbers has a very wide range of application.

One point should be made clear here. If the numbering of the population is carried out in such a way as to be independent of certain characteristics of the population, any set of numbers will serve to draw a sample random with respect to those characteristics. The randomness in such a case lies, so to speak, in the allocation of ordinals to the population, not in deciding which ordinals to select for the sample. But in practice a procedure of this kind is of no value, since it only throws back to the difficulty of numbering the population "at random". The usual course is to number the population in any convenient way, related to the characteristics or not, and then seek for a set of numbers which are a random set from the possible ordinals of the population.

8.11. One of the more obvious ways of drawing random samples from an enumerated population is to use haphazard numbers taken from some totally unrelated source. Suppose, for instance, we wished to take a sample from the visible stars in the sky. We will ignore the small complications due to the existence of double stars and unresolved objects. Since the position of a star on the celestial sphere is defined by latitude and longitude, what is then required is a series of random pairs of latitudes and longitudes. At first sight it seems plausible to take an ordinary atlas and choose the figures set out in the index for place-names arranged alphabetically; for there is little reason to expect any relationship between the

distribution of stars in the sky and the distribution of places on the Earth's surface. A little reflection, however, will show that the method is unsound. There are large stretches of territory and sea on the Earth which have no place-names on them—the poles, deserts and oceans; consequently no numbers will occur for these regions and there will be corresponding areas on the celestial sphere which have no chance of being included.

8.12. As a next attempt we might take a book containing a number of digits, e.g. a telephone directory, or a set of statistical tables or mathematical tables, open it at hazard and choose the digits which first strike the eye, or which occur at the top of the page, and so on. This is an improvement, but it is still open to some objection.

(*a*) Telephone directories. Table 1.4 on page 6 shows the distribution of 10,000 digits taken from the London telephone directory. Pages were chosen by opening the directory haphazardly, numbers of less than four digits and numbers in heavy type were ignored; and of the four-figure numbers remaining the two right-hand ones were taken for all numbers on the page. If the numbers were random we should expect about 1000 of each digit in the total of 10,000. Actually there are very considerable deviations from this expectation, and we shall see in a later chapter that they cannot be explained as sampling fluctuations. There are significant deficiencies in 5's and 9's, due to several causes such as the tendency to avoid these digits because they sound alike, the reservation of numbers ending in 99 for testing purposes by telephone engineers and so on. It is evident that tables of random numbers could not be constructed from directories such as this.

(*b*) Mathematical tables. Evidently care has to be exercised in using mathematical tables in constructing random series. Suppose, for instance, we take a set of logarithm tables. There are clearly relationships between successive logarithms, expressible by the fact that differences are approximately constant if the interval is small. Moreover there is a very curious theorem about digits in certain classes of table which throws theoretical doubt on the method. Consider the logarithms to base 10 of the natural numbers from 1 onwards. Suppose we choose the kth digit in each and so obtain a series of numbers 0–9. Then the proportional frequency of any digit in this series does *not* tend to a limit as the length of the series increases, whatever k may be.* Just what does happen does not appear to be known, but it would seem that certain systematic effects begin to show themselves and these will obviously endanger the randomness of the series.

(*c*) Statistical tables. If we have a volume of statistics such as populations of towns and rural districts there are some grounds for supposing that if the numbers are large—say four figures or more—the final digits will be random. Here again, however, the use of such tables requires care—they may have been compiled by an observer with number preferences, and some rounding up may have taken place.

8.13. However, the necessity for the ordinary student to construct random series of his own has been obviated by the publication of various tables of Random Sampling Numbers. There are three such available:—

(*a*) Tippett's numbers comprise 41,600 digits taken from census reports combined into fours to make 10,400 four-figure numbers (*Tracts for Computers*, No. 15).

(*b*) Kendall and Babington Smith's numbers comprise 100,000 digits grouped in twos and fours and in 100 separate thousands (*Tracts for Computers*, No. 24). These numbers were

* Cf. J. Fanel, *Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich* (1917), **62**, 286. So great a mathematician as Poincaré made a mistake on this point.

obtained from a machine specially constructed for the purpose on the lines very briefly described in Example 8.3.

(c) Fisher and Yates' numbers comprise 15,000 digits arranged in twos (*Statistical Tables for Biological, Agricultural and Medical Research*). These numbers were obtained from the 15th–19th digits in A. J. Thompson's tables of logarithms and were subsequently adjusted, it having been found that there were too many sixes.

Before considering the basis of these tables it may be helpful to give some examples of their use. Here are the first 200 of the Kendall–Babington Smith tables :—

TABLE 8.5

Random Sampling Numbers.

(*Tracts for Computers*, No. 24.)

23 15	75 48	59 01	83 72	59 93	76 24	97 08	86 95	23 03	67 44
05 54	55 50	43 10	53 74	35 08	90 61	18 37	44 10	96 22	13 43
14 87	16 03	50 32	40 43	62 23	50 05	10 03	22 11	54 38	08 34
38 97	67 49	51 94	05 17	58 53	78 80	59 01	94 32	42 87	16 95
97 31	26 17	18 99	75 53	08 70	94 25	12 58	41 54	88 21	05 13

Example 8.5

To draw a sample of 10 men from the population of 8585 men of Table 1.7.

The first process is to number the population ; and here, as in most similar cases, one numbering has already been provided by the frequency-distribution. We take numbers 1 and 2 to be those in the group 57– inches, numbers 3 to 6 those in the group 58–, and so on, those in the group 77– inches being numbers 8584 and 8585.

Now we take 10 four-figure numbers from the tables, e.g. reading across in Table 8.5 we have

2315, 7548, 5901, 8372, 5993, 7624, [9708], [8695], 2303, 6744, 0554, 5550.

The two numbers in square brackets are greater than 8585 and we ignore them. We now select the individuals corresponding to the remaining 10 numbers. They will be found to be in the intervals 65–, 70–, 68–, 72–, 68–, 70–, 65–, 69–, 63–, 68– inches respectively.

The mean of these values considered as located at the centres of intervals is 68·24, as against a value in the population of 67·46.

Example 8.6

To draw a sample of 12 from the population in the following bivariate table, showing the relation between inoculation and attack in cholera.

	Not Attacked.	Attacked.	Total.
Inoculated	276 (0001–3312)	3 (3313–3348)	279
Not inoculated . .	473 (3349–9024)	66 (9025–9816)	539
Totals	749	69	818

There are now 818 members. We could, of course, take three-figure numbers from the tables, obtaining, e.g. from Table 8.5

231, 575, 485, etc.

But this is rather troublesome as the numbers are not grouped in threes. It is more convenient to take four-figure numbers as before and to associate each member of the population with 12 numbers in the tables, e.g. the first would correspond to 0000–0011, the second to 0012–0024, and so on. We then get the numbers shown in brackets in the above table. Numbers above 9816 we ignore as before.

The two numbers omitted in the previous example can now be used, and we find the following results :—

	Not Attacked.	Attacked.	TOTAL.
Inoculated	3	0	3
Not inoculated . .	8	1	9
TOTALS	11	1	12

Here, for example, the member corresponding to the number 2315 falls in the not-attacked : inoculated class, and so on.

It has so happened in this example that no member in the very small class inoculated : attacked class has been selected. Suppose we had had a series containing

3314, 3323, 3333, 3341.

All these fall into the group and there are four of them, as against only three members in the population. Had we been confronted with this position we should have had to decide whether the sampling was to be with or without replacement. If it was without replacement, we should have to suppose that the first three numbers in the group 3313–3348 exhausted that part of the population and ignore all numbers of the group occurring subsequently.

Example 8.7

To construct a series of random permutations of the numbers 1 to 5.

Here we are not concerned with the digits 0, 6, 7, 8 and 9 and so ignore them in the table of random numbers. We read through the table and note the digits as they occur, e.g. in Table 8.5 we have 2315, 7548, etc. The 7 is to be ignored and also the second 5, for one 5 has already occurred. We then reach the permutation 23154. Then we start again, the next series being 8, 5901, 8372, 5993, 7624, etc., giving the permutation 13254 ; and so on.

Example 8.8

To take a random sample of 10 from the normal population $dF = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

This is a particularly interesting case, for we have to select a sample from an infinite population. Such a process, as has been seen, can only be considered as a limiting one.

Consider the frequencies of the normal curve in ranges 0·1 on each side of the mean. These may be obtained very simply from tables of the normal integral by differencing and in fact are given in many tables of that integral, e.g. that of Appendix Table 2. Suppose the frequencies rounded up to four places of decimals, e.g. those near the mean would be

0·0–	0·0398
0·1–	0·0394
0·2–	0·0387
0·3–	0·0375, etc.

and the total frequencies are given by the normal integral itself, e.g.

Upper Limit of Interval.	Frequency up to that Limit.	Upper Limit of Interval.	Frequency up to that Limit.
0·0	0·5000	− 0·1	0·4602
0·1	0·5398	− 0·2	0·4207
0·2	0·5793	− 0·3	0·3821
0·3	0·6179, etc.	− 0·4	0·3085, etc.

We may now attach a four-figure random number to this population, which is finite and discontinuous : e.g. the number 5461 corresponds to a variate-value + 0·1– and the number 3500 to − 0·4–.

Had we taken the table to n places of decimals we should have required n– figure numbers. Furthermore, we can make the approximation more exact by taking a finer variate interval. Such matters as this are to be decided in the light of the degree of approximation required.

8.14. Random Sampling Numbers must obey certain conditions before they can be used. Any set of numbers whatever is random in the sense that it might arise, with however great improbability, from random sampling ; but such a set might not be suitable as a table of Random Sampling Numbers. From the examples already given it is clear that we desire such a table to have very great flexibility. It should give random results in as many cases as possible, whether used in part or in whole.

Now it is impossible to construct a table of Random Sampling Numbers which will satisfy this requirement entirely. Suppose, to take an extreme case, we constructed a table of $10^{10^{10}}$ digits. The chance of any digit being a zero is $\frac{1}{10}$ and thus the chance that any given block of a million digits are all zeros is 10^{-10^6}. Such a set should therefore arise fairly often in the set of $10^{(10^{10}-6)}$ blocks of a million. If it did not, the whole set would not be satisfactory for certain sampling experiments. Clearly, however, the set of a million zeros is not suitable for drawing samples in an experiment requiring less than a million digits.

Thus, it is to be expected that in a table of Random Sampling Numbers there will occur patches which are not suitable for use by themselves. The unusual must be given a chance of occurring in its due proportion, however small. Kendall and Babington Smith attempted to deal with this problem by indicating the portions of their table (5 thousands out of 100) which it would be better to avoid in sampling experiments requiring fewer than 1000 digits.

8.15. If a table of random numbers is used to draw members from a population of ten, we expect the members to appear in approximately equal proportions. In other words we expect such a table to contain the ten digits 0–9 in approximately equal pro-

portions. Similarly we expect the hundred pairs 00–99 to appear in approximately equal proportions, and so on. Various tests of this kind, based on a comparison between actual frequencies and those required to satisfy the laws of probability, can be devised. No table can satisfy them all, but if it satisfies tests which (*a*) ensure the randomness of the numbers for the commoner types of sampling inquiry for which it is likely to be used and (*b*) are capable of revealing any particular sort of bias to which the numbers are susceptible in virtue of their mode of formation, it is likely to be of general application.

For a more detailed discussion of these matters and the results of tests on the Tippett tables, the Kendall–Babington Smith tables and the Fisher–Yates tables, reference may be made to the works listed at the end of the chapter.

Sampling from a Continuous Population

8.16. Random Sampling Numbers offer the best method known at the present time of drawing random samples from an enumerable universe, and as was seen in Example 8.6, may also be used to draw samples from a continuous population specified mathematically. But cases sometimes occur in which they cannot be employed. For instance, if we wish to take a sample of milk or flour, we cannot in practice number each particle and extract it from the population for examination. In such cases we are usually compelled to fall back on more intuitively grounded procedure. To take a random sample from a milk churn, for example, we might stir the contents thoroughly and scoop up a sample haphazardly. Sometimes, when the population is of manageable size, we can proceed systematically by dividing it into a number of parcels and selecting parcels by the ordinary technique of random numbers. Most sciences have their own peculiar sampling problems and no attempt can be made here to discuss them all. At this point we leave the technique of random sampling and assume hereafter, unless the contrary is stated, that the material we are discussing has been obtained by a random process.

Sampling from Attributes

8.17. As an introduction to the general sampling problems we shall consider the sampling of attributes, which raises all the difficulties of principle but is not obscured by too much mathematics.

Suppose we have a random sample from a population whose members all exhibit either an attribute A or its negative not-A. Our sample is n in number, and a proportion p, or a number pn, exhibit the attribute; and consequently a proportion q, or number qn $(p + q = 1)$ do not. We will assume that the population is large, or that sampling is with replacement, so that the probability of obtaining an A at any drawing is not affected by other drawings and is therefore a constant, say ϖ.

The problems we have to consider are of three types :—

(*a*) Suppose we have some reason for supposing that the proportion of A's in the population is given by a known ϖ. Does the observed proportion p bear out this hypothesis or is it so divergent from ϖ as to lead us to doubt the hypothesis? In an experiment with plants exhibiting two strains of a quality such as height in pea plants, we may wish to test whether the breeding follows the simple Mendelian law of dominant and recessive. If we begin with two pure strains tall and short, cross-breed a first generation and then produce a second generation by interbreeding, the proportional frequencies of " short " and " tall " in this generation will be $\frac{3}{4}$ and $\frac{1}{4}$ if " short " is dominant and $\frac{1}{4}$ and $\frac{3}{4}$ if " tall " is dominant, provided that the simple Mendelian law holds. Suppose we carry out such

The method of Bayes will give the same result if we suppose the possible values of ϖ equally distributed between 0 and 1 for $d\varpi \to 0$. For then

$$P(\varpi|p) = \frac{\binom{n}{np}\varpi^{np}\chi^{nq}}{\binom{n}{np}\varpi^{np}\chi^{nq}} \tag{8.3}$$

$$= \binom{n}{np}\varpi^{np}\chi^{nq} \tag{8.4}$$

which, as before, is maximised when $\varpi = p$.

There is another way of looking at this problem of estimation. Suppose we took a large number of samples from the population of ϖ A's and χ not-A's. Our estimate of ϖ would be p in each case, p varying from sample to sample; and the mean value of all such estimates would be

$$\begin{aligned} E(p) &= \sum_{p=0}^{n} p\binom{n}{np}\varpi^{np}\chi^{nq} \qquad (8.5) \\ &= \frac{1}{n}\Sigma\left\{np\binom{n}{np}\varpi^{np}(\chi)^{nq}\right\} \\ &= \frac{1}{n}\varpi . n\Sigma\binom{n-1}{np-1}\varpi^{np-1}(\chi)^{nq} \\ &= \varpi\{\varpi + \chi\}^{n-1} \\ &= \varpi, \end{aligned}$$

so that the mean value of our estimate over all possible samples is ϖ. Such an estimate is called *unbiased*—if we follow the rule of estimation the average of our estimates in a large number of cases will be exactly the correct value ϖ. It may thus be argued that the unbiased estimate should be taken as a reliable estimate of ϖ.

8.22. In this case, therefore, all the approaches lead to the same conclusion (a happy state of affairs which, as we shall see in the sequel, does not always exist). Consider now the next stage of the problem: what is the reliability of the estimate? In other words, how far is the estimate likely to differ from the true value?

We know that if the sample value p differs from ϖ by $t\sqrt{\frac{\varpi\chi}{n}}$, the probability of the difference becomes smaller as t increases. Thus, with an assigned degree of probability we can say that it is improbable that p will differ from ϖ by more than an assigned amount. But to specify this amount exactly we require to know ϖ; and this is precisely the quantity we are trying to find.

The problem can only be solved as an approximation. If n is large the standard error of ϖ is of the order $n^{-\frac{1}{2}}$, so that we may put

$$\varpi = p + \frac{k}{n^{\frac{1}{2}}}.$$

Thus

$$\begin{aligned} \sqrt{\frac{\varpi\chi}{n}} &= \sqrt{\left\{\frac{1}{n}\left(p + \frac{k}{n^{\frac{1}{2}}}\right)\left(q - \frac{k}{n^{\frac{1}{2}}}\right)\right\}} \\ &= \sqrt{\left\{\frac{pq}{n}\left(1 + \frac{k(q-p)}{pqn^{\frac{1}{2}}}\right)\right\}}; \end{aligned}$$

neglecting terms of order n^{-2},

$$= \sqrt{\frac{pq}{n}}\left\{1 + \frac{1}{2}\frac{k(q-p)}{pqn^{\frac{1}{2}}}\right\}. \quad . \quad . \quad . \quad . \quad . \quad (8.6)$$

Thus for large n the standard error of ϖ is approximately equal to $\sqrt{\frac{pq}{n}}$; and we thus reach the fundamental result that in large samples of attributes the standard error may be calculated by using the estimates of the parameters under estimate instead of the (unknown) values of those parameters themselves.

Example 8.11

In a sample of 600, 240 are found to possess the attribute A. Thus $p = 0{\cdot}40$, $np = 240$, $\sqrt{(npq)} = 12$. We can thus regard it as somewhat improbable that $n\varpi$ differs from 240 by more than twice this amount, 24, and highly improbable that it differs by more than 36. We thus can say with some confidence that $n\varpi$ lies in the range 240 ± 24 and with great confidence that it lies in the range 240 ± 36.

8.23. We now turn to a general consideration of the problems of sampling which have been exemplified above. In the first place, let us note the role of the sampling distribution in this branch of the subject. We construct from the observations some statistic t. The sampling distribution of this statistic will in general (but not always) depend on some parameters of the parent population. The probability of the observed t then permits the making of statements, by inverse probability; likelihood or otherwise, about these parameters, and thus we are enabled to draw inferences about the parent population. The sampling distribution is thus fundamental to the whole subject and several subsequent chapters will be devoted entirely to the methods of finding distributions when the parent is specified.

If we wish to test some hypothesis about the parent which is expressible by the determination of certain parameters *a priori*, the problem is fairly simple. Given the values of the parameters, we can determine from the sampling distribution the probability of the observed value of the statistic, and use this to assess the acceptability of the hypothesis. Complications can arise even here, however, for in general, several statistics can be compiled from the same sample, and they need not necessarily all lead to the same conclusion about the hypothesis; for example, a sample might have a mean which throws doubt on the hypothesis and a variance which does not. We shall discuss this difficulty more fully in the second volume.

8.24. When the parameters of the population are not given *a priori*, we have the double problem of estimating the parameters from the sample and assigning probable limits to the estimates so obtained. We have already touched on some of the principles of estimation and shall develop the topic more systematically in due course. When we have obtained an estimate—itself a statistic—we seek its sampling distribution and therefrom can assign probable limits to the population value. A special class of cases arises when we can find a statistic whose sampling distribution depends on only one parameter of the population (as in the case of attributes).

8.25. These latter types of problem permit of certain important approximations, namely in the case when the sample is large. We saw in Chapter 7 that under very general conditions the sum of n independent variables, distributed in whatever form, tends to normality as n tends to infinity. Now many of the ordinary statistics in current use can

be expressed as the sum of variates, e.g. all the moments; and many others may also be shown to tend to normality for large samples. Thus we may approximate—

(*a*) By taking a statistic, calculated from the sample as if it were a population, to be the estimate of the corresponding statistic in that population, e.g. the variance of the sample may be taken as an estimate of the variance of the population.

(*b*) By calculating the mean and variance of the sampling distribution by using, instead of the unknown parameter values, the statistic values calculated according to (*a*).

(*c*) By assuming that the distribution is normal and hence determining probabilities from the normal integral with the aid of the sampling mean and sampling variance (the latter being the square of the standard error).

8.26. Just how large n must be for such approximations to be valid is not always easy to say. For some distributions, particularly that of the mean, quite a satisfactory approximation is given by low values of n, say $n > 30$. For others n has to be much higher before the approximation begins to give satisfactory results, e.g. for the product-moment correlation coefficient (below, **14.5**) even values as high as 500 are not good enough.

8.27. In the following three chapters we discuss the approximate and accurate methods for determining sampling distributions. Chapter 9 deals with large samples and is thus devoted mainly to methods for determining standard errors. Chapter 10 deals with methods for determining sampling distributions exactly. Chapter **11** discusses methods of approximating to sampling distributions by finding their lower moments.

NOTES AND REFERENCES

For some interesting discussions of problems of sampling generally, see Jensen (1926), Bowley (1926), Hilton (1924), Kiser (1934), Yates (1935) and Neyman (1934). For a discussion of random sampling, see Kendall and Babington Smith (1938 and 1939) and Kendall (1941). The various tables of random numbers are referred to in the Introduction.

Bowley, A. L. (1926), "Measurement of the Precision attained in Sampling," *Bull. Int. Stat. Inst.*, **22**, premier livre.

Hilton, John (1924), "Enquiry by sample; an experiment and its results," *Jour. Roy. Statist. Soc.*, **87**, 544.

Jensen, A. (1926), "Report on the representative method in statistics," *Bull. Int. Stat. Inst.*, **22**, premier livre.

Kendall, M. G., and Babington Smith, B. (1938), "Randomness and random sampling numbers," *Jour. Roy. Statist. Soc.*, **101**, 147, and (1939) "Second paper on random sampling numbers," *Supp. J.R. Statist. Soc.*, **6**, 51.

Kendall, M. G. (1941), "A theory of randomness," *Biometrika*, **32**, 1.

Kiser, C. V. (1934), "Pitfalls in sampling for population study," *Jour. Amer. Stat. Assoc.*, **29**, 250.

Neyman, J. (1934), "On two different aspects of the representative method," *Jour. Roy. Statist. Soc.*, **97**, 558.

Yates, F. (1935), "Some examples of biased sampling," *Ann. Eugen., Lond.*, **6**, 202.

Yule, G. Udny (1927), "On reading a scale," *Jour. Roy. Statist. Soc.*, **90**, 570.

EXERCISES

8.1. Of 10,000 babies born in a particular country 5100 are male. Taking this to be

a random sample of the births in that country, show that it throws considerable doubt on the hypothesis that the sexes are born in equal proportions.

Consider how far this conclusion would be modified if the sample consisted of 1000 births, 510 of which were male.

8.2. If the number of members of a population bearing an attribute A is relatively small, show that the standard error of the number of A's in the sample is the square root of that number. Show also that the number of A's in the sample is an unbiased and a maximum likelihood estimate of the parameter of the Poisson distribution expressing the distribution of the number of A's in large samples from the population.

8.3. By considering the hypergeometric distribution, show that if samples of n are drawn from a finite population of N without replacement, and a proportion ϖ of that population bear an attribute A, then the standard error of the proportion p in the sample is

$$\left(\frac{N-n}{N-1}n\varpi\chi\right)^{\frac{1}{2}}.$$

Show also that p is an unbiased estimate of ϖ.

8.4. (Tchebycheff's inequality). Show that for any distribution

$$\int_{-\infty}^{\infty}\left(1-\frac{x^2}{t^2}\right)dF \leqslant \int_{-t}^{t} dF$$

and hence that for any member drawn at random

$$P(|\,x\,| > \alpha\sqrt{\mu_2}) \leqslant \frac{1}{\alpha^2}.$$

Show further that the variance of the sampling distribution of proportions bearing an attribute A in samples of n from a population of attributes is not greater than $\frac{1}{4n}$. Hence the probability that an observed proportion p differs from the true proportion ϖ by more than amount k is not greater than $\frac{1}{4nk^2}$.

(This gives us an exact result, no assumptions about the normality of the limiting form of the binomial or the use of estimates in calculating standard errors being involved. The limits are, however, much too wide.)

8.5. If a proportion ϖ has to be estimated from a simple random sample with proportion p, and if f is the prior probability of ϖ, then the posterior probability of ϖ is, according to Bayes' theorem, proportional to

$$f\varpi^{np}(1-\varpi)^{nq}.$$

Show that this is a maximum if

$$\frac{1}{f}\frac{\partial f}{\partial\varpi} + n\frac{p(1-p)}{\varpi(1-\varpi)} = 0.$$

Hence, in general, as n increases, the solution tends to $\varpi = p$, whatever the prior probability of ϖ. In other words, the maximum likelihood estimate is an approximation to that given by Bayes' theorem as n tends to infinity, even if Bayes' postulate is not assumed.

CHAPTER 9

STANDARD ERRORS

9.1. Towards the close of the last chapter we discussed the estimation of statistical parameters from large samples and the type of judgment of their reliability which depends on the use of the standard error. It was remarked that, for large samples, an estimate of a parameter may be obtained by calculating from the sample values the value of the parameter in the sub-population composed by the sample; and it was established that for samples of n the standard error gives a valid measure of precision, provided that (a) the sampling distribution of the statistic under discussion approaches normality and (b) that n is large in the sense there defined. It was also pointed out that a sufficiently accurate estimate of standard errors involving parent parameters could be obtained by using as the parameter values the corresponding statistics from the sample itself.

Since the majority of statistics in current use do tend to normality the theory of large samples is, in the main, devoted to the determination of standard errors. In this chapter we describe the principal methods available for the purpose, and incidentally derive formulae for the standard errors of the various statistics considered in previous chapters. To avoid the usual square roots associated with the standard error we shall write our results as sampling variances and covariances. Thus, for a statistic t we write the variance of its sampling distribution as var t. The covariance of the joint distribution of two statistics t and u, that is, the first product-moment of their joint sampling distribution, is written cov (t, u). We shall also consider the distributions in large samples of some statistics which do not tend to normality.

9.2. By definition, the rth moment of a statistic t, that is the rth moment of its sampling distribution, is the mean value of t^r taken over all possible samples, and may be written $E(t^r)$ (cf. **3.35**). If the joint distribution of the variates $x_1 \ldots x_n$, from which t is calculated, is $dF(x_1 \ldots x_n)$, then the rth moment of t is the integral of $t^r dF$ (considered as a function of the x's) over the domain of the x's. In particular, if the sample is simple and random and the parent distribution is dF, we have

$$E(t^r) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} t^r \, dF(x_1) \qquad dF(x_n)$$

We are particularly interested in this chapter in the first and second moments of t, that is, the mean and variance of its sampling distribution. It may be recalled that the mean value of a sum is the sum of the mean values and that, if the variables are independent, the mean value of a product is the product of the mean values (**3.36**). These two results will be repeatedly required.

Standard Errors of Moments

9.3. In the first place we consider the standard errors of the wide class of statistics depending on the moments, including the mean, variance, the Pearson measures of skewness and kurtosis and the moments and cumulants themselves.

The sampling distributions of moments tend to normality under very general conditions in virtue of the Central Limit Theorem. In fact, if the parent distribution is represented

by $f(x)\,dx$, the distribution of the jth power of x, say y, is easily seen to be $\frac{1}{j}f\left(y^{\frac{1}{j}}\right)y^{\frac{1-j}{j}}\,dy$ and the jth moment is the sum of n independent variates, each of which is distributed in that form.

It is not so obvious that functions of the moments such as b_1 and b_2 (the sample values of the parameters β_1 and β_2) will tend to normality, and special investigations may have to be made for particular statistics. Even at the present time it is often assumed without proof that certain statistics tend to normality, the feeling apparently being (so far as any feeling uprises into consciousness) that as most statistics do tend to normality the onus is on an objector to prove that any particular statistic does not. This is very dangerous to accurate inferential reasoning and the point is one to be borne in mind wherever a standard error is used.

On a similar point, it should also be remembered that some statistics tend to normality more rapidly than others, and a given n may be large for some purposes but not for others. So far as it is possible to generalise with safety, we can usually (but not always) assume values of n greater than 500 to be large; values greater than 100 are often great enough to be large for our purposes; values below 100 are suspect in many instances; and values below 30 are very rarely large.

In the following we shall adopt the usual convention in regard to the distinction of parameters and statistics by writing Greek letters to represent the former and Roman letters to represent the latter. We have, then, for the rth moment-statistic m'_r, corresponding to the rth moment parameter μ'_r,

$$m'_r = \frac{1}{n}\sum_{j=1}^{n}(x_j^r) \qquad (9.1)$$

and for the mean-moment

$$m_r = \frac{1}{n}\sum_{j=1}^{n}(x_j - m'_1)^r. \qquad (9.2)$$

9.4. Consider now the mean value of m'_r. Since the x's are independent we have

$$\begin{aligned} E(m'_r) &= \frac{1}{n}\Sigma E(x^r) \\ &= E(x^r) \\ &= \mu'_r. \end{aligned} \qquad (9.3)$$

The sampling variance of m'_r is, by definition, $E(m'_r - \mu'_r)^2$ and thus

$$\begin{aligned} \text{var}\,(m'_r) &= E\left\{\frac{1}{n}\Sigma(x^r) - \mu'_r\right\}^2 \\ &= \frac{1}{n^2}E[\{\Sigma(x^r)\}^2 - 2n\mu'_r\Sigma(x^r) + n^2\mu'^2_r] \\ &= \frac{1}{n^2}E[\{\Sigma(x^r)\}^2] - \mu'^2_r \\ &= \frac{1}{n^2}E\{\Sigma(x_j^{2r}) + \Sigma(x_j^r x_k^r)\} - \mu'^2_r, \end{aligned}$$

the second summation extending over the $n(n-1)$ cases in which $j \neq k$ (permutations

of j and k thus being allowed). Since the x's are independent the mean value of the product is the product of the mean values, and thus

$$\begin{aligned}\text{var}\,(m'_r) &= \frac{1}{n^2}\{n\mu'_{2r} + n(n-1)\mu'^2_r\} - \mu'^2_r \\ &= \frac{1}{n}(\mu'_{2r} - \mu'^2_r). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.4)\end{aligned}$$

This is an exact result.

In a similar way, if we have two moments, m'_q, m'_r, their sampling covariance is given by

$$\begin{aligned}\text{cov}\,(m'_q, m'_r) &= E\{(m'_q - \mu'_q)(m'_r - \mu'_r)\} \\ &= E\left\{\left(\frac{1}{n}\Sigma x^q - \mu'_q\right)\left(\frac{1}{n}\Sigma x^r - \mu'_r\right)\right\} \\ &= \frac{1}{n^2}E(\Sigma x^{q+r}) - \frac{1}{n}\mu'_q E(\Sigma x^r) - \frac{1}{n}\mu'_r E(\Sigma x^q) + E(\mu'_q\mu'_r) \\ &= \frac{1}{n}(\mu'_{q+r} - \mu'_q\mu'_r), \qquad (9.5)\end{aligned}$$

which reduces to (9.4) if $q = r$, as it must, for the first product-moment of two identical variables is their variance.

9.5. The formulae for moments about the mean are not so simple, for the mean itself is subject to sampling fluctuations. We have, in fact,

$$E(m_r) = \frac{1}{n}E\{\Sigma(x - m'_1)^r\}. \qquad (9.6)$$

Now putting $r = 1$ in (9.4) we find

$$\begin{aligned}\text{var}\,(m'_1) &= \frac{1}{n}(\mu'_2 - \mu'^2_1) \\ &= \frac{1}{n}\mu_2 \qquad (9.7)\end{aligned}$$

and thus the standard error of the mean is $\sqrt{\dfrac{\mu_2}{n}}$. Consequently, if the distribution is anywhere near normality, nearly all the values of x will lie within a range of the true value of order $n^{-\frac{1}{2}}$. To order n^{-1} we may then, taking an origin at the mean of the parent population, neglect powers of m'_1 higher than the first, and we then obtain from (9.6)

$$\begin{aligned}E(m_r) &= \frac{1}{n}E\{\Sigma(x^r - rm'_1x^{r-1})\} \\ &= \frac{1}{n}E\left\{\Sigma x^r - r\frac{1}{n}\Sigma x\Sigma x^{r-1}\right. \\ &= \frac{1}{n}E\left\{\left(1 - \frac{r}{n}\right)\Sigma x_j^r - \frac{r}{n}\sum_{j,k}(x_jx_k^{r-1})\right\}, \qquad j \neq k.\end{aligned}$$

Now the second term in the expectation on the right will involve the moments $\mu'_1\mu'_{r-1}$ and will vanish since we have chosen our origin at the mean of the parent distribution ($\mu'_1 = 0$). We shall then have, to order n^{-1}

$$E(m_r) = \mu_r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.8)$$

a result which is not, like (9.3), exact, but is an approximation to order n^{-1}. To this order we have

$$\begin{aligned}\text{var}(m_r) &= \frac{1}{n^2}E(m_r - \mu_r)^2 \\ &= \frac{1}{n^2}E(m_r^2) - \mu_r^2 \\ &= \frac{1}{n^2}E\left\{\Sigma(x_j^r) - \frac{r}{n}\Sigma(x_j x_k^{r-1})\right\}^2 - \mu_r^2 \\ &= \frac{1}{n^2}E\left\{\Sigma(x_j^{2r}) + \Sigma(x_j^r\, x_k^r) + \frac{r^2}{n^2}\Sigma(x_j^2\, x_k^{2r-2})\right. \\ &\quad \left. + \frac{r^2}{n^2}\Sigma(x_j^2\, x_k^{r-1}\, x_l^{r-1}) - \frac{2r}{n}\Sigma(x_j^{r+1}\, x_k^{r-1})\right\} - \mu_r^2 \quad j \neq k \neq l.\end{aligned}$$

The expectations of other terms occurring in the squaring vanish, since they contain μ_1. The expectation of $\frac{1}{n^2}.\frac{r^2}{n^2}\Sigma(x_j^2\, x_k^{2r-2})$ is of order $\frac{1}{n^2}$ and is thus to be neglected. The remaining terms give us

$$\text{var}(m_r) = \frac{1}{n}(\mu_{2r} - \mu_r^2 + r^2\mu_2\,\mu^2_{r-1} - 2r\mu_{r-1}\,\mu_{r+1}). \tag{9.9}$$

Similarly it appears that

$$\text{cov}(m_r, m_q) = \frac{1}{n}(\mu_{r+q} - \mu_r\mu_q + rq\mu_2\,\mu_{r-1}\,\mu_{q-1} - r\mu_{r-1}\,\mu_{q+1} - q\mu_{r+1}\,\mu_{q-1}) \ . \tag{9.10}$$

Example 9.1

From (9.7) we have

$$\text{var}(m_1') = \frac{\mu_2}{n}.$$

Now, for the height distribution of Table 1.7 we found (Examples 2.1 and 2.6) that $m_1' = 67{\cdot}46$ $\sqrt{m_2} = 2{\cdot}57$. Suppose we regard this distribution as a simple random sample from the adult male inhabitants of the United Kingdom living at the time when the data were collected. What can we say about the mean of the population ?

The standard error of the mean depends on μ_2. This is an unknown quantity, but we may, in accordance with the general principles of large sample theory, use m_2 instead. We then find

$$\text{Standard error of } m_1' = \sqrt{\frac{2{\cdot}57}{8585}} = 0{\cdot}028 \text{ approximately.}$$

Thus we can say that the population mean probably lies in the range of twice this amount on either side of the sample mean, i.e. in the range $67{\cdot}46 \pm 0{\cdot}056$, and very probably in thrice the range, i.e. $67{\cdot}46 \pm 0{\cdot}084$. Our estimate of the mean would almost certainly be less than a tenth of an inch in error.

Example 9.2

From equation (9.9) with $r = 4$ we find

$$\text{var}(m_4) = \frac{1}{n}(\mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_2\mu_3^2).$$

In Chapter 11 we shall show how to obtain this result by other methods of a more exact character and confirm that it is, in fact, exact to order n^{-1}.

Example 9.3

To show that in samples from a symmetrical population the first product–moment between the mean and any mean–moment of even order vanishes to order n^{-1}.

We have, by definition

$$\text{cov}\,(m_1', m_r) = \frac{1}{n^2}E\left[\left\{\Sigma(x)\{\Sigma(x^r) - \frac{r}{n}\Sigma(x_j x_k^{r-1})\}\right\}\right]$$

$$= \frac{1}{n^2}E\left\{\Sigma(x^{r+1}) - \frac{r}{n}\Sigma(x_j^2\, x_k^{r-1})\right\},$$

the other terms vanishing, since they involve the unit power of x, if we take an origin at the mean of the parent population,

$$= \frac{1}{n}(\mu_{r+1} - r\mu_2\mu_{r-1}).$$

Now if r is even, μ_{r+1} and μ_{r-1}, being moments of odd order, will vanish for a symmetrical population and hence

$$\text{cov}\,(m_1', m_r) = 0.$$

In the language of the theory of correlation (Chapter 14) the mean and the even moment about the mean are uncorrelated to order n^{-1}.

Standard Errors of Functions of Moments

9.6. From the expressions we have just derived for the sampling variances and covariances of moments, approximate expressions can be obtained for the sampling variances of functions of the moments. Suppose $\phi(m)$ is such a function. We have the functional relation in differences

$$\Delta\phi = \frac{\partial\phi}{\partial m_1}\Delta m_1 + \frac{\partial\phi}{\partial m_2}\Delta m_2 + \ . \ . \ . + O(\Delta m)^2. \qquad (9.11)$$

Now any variations in m due to fluctuations of sampling are of order $n^{-\frac{1}{2}}$. To our approximation, therefore, we may neglect the terms of order $(\Delta m)^2$ in (9.11), and the variation $\Delta\phi$ is then seen to be a linear function of the variations Δm and is equivalent to an equation in differentials; that is to say, since the m's are distributed normally in the limit, so will ϕ be. We have, from (9.11),

$$E(\phi) = \Sigma\frac{\partial\phi}{\partial m_j}E(m_j).$$

Hence, measuring from the means of the m's and ϕ, we find, squaring and taking mean values,

$$\text{var}\,(\phi) = \Sigma\left\{\left(\frac{\partial\phi}{\partial m_j}\right)^2 \text{var}\,(m_j)\right\} + \Sigma\left\{\frac{\partial\phi}{\partial m_j}\frac{\partial\phi}{\partial m_k}\,\text{cov}\,(m_j, m_k)\right\} \qquad (9.12)$$

the first summation extending over all the m's appearing in ϕ, and the second over all m's such that $j \neq k$.

Similarly, for two functions ϕ_1, ϕ_2, we have

$$\text{cov}\,(\phi_1, \phi_2) = \Sigma\left\{\frac{\partial\phi_1}{\partial m_j}\frac{\partial\phi_2}{\partial m_j}\,\text{var}\,(m_j)\right\} + \Sigma\left\{\frac{\partial\phi_1}{\partial m_j}\frac{\partial\phi_2}{\partial m_k}\text{cov}\,(m_j, m_k\right. \qquad (9.13)$$

Example 9.4

To find the sampling variance of the fourth cumulant. We have

$$\kappa_4 = \mu_4 - 3\mu_2^2$$
$$d\kappa_4 = d\mu_4 - 6\mu_2 d\mu_2.$$

Hence, squaring and taking mean values,

$$\text{var}(\kappa_4) = \text{var}(\mu_4) - 12\mu_2 \text{cov}(\mu_4, \mu_2) + 36\mu_2^2 \text{var}(\mu_2).$$

Making the appropriate substitutions from (9.9) and (9.10) we have

$$\text{var}(\kappa_4) = \frac{1}{n}\{\mu_8 - \mu_4^2 + 16\mu_2\mu^2\mu_3^2 - 8\mu_3\mu_5 - 12\mu_2(\mu_6 - \mu_4\mu_2 - 4\mu_3^2) + 36\mu_2^2(\mu_4 - \mu_2^2)\}$$
$$= \frac{1}{n}\{\mu_8 - 12\mu_6\mu_2 - 8\mu_5\mu_3 - \mu_4^2 + 48\mu_4\mu_2^2 + 64\mu_3^2\mu_2 - 36\mu_2^4\}.$$

For a normal parent, $\mu_4 = 3\sigma^4$, $\mu_6 = 15\sigma^6$, $\mu_8 = 105\sigma^8$ and we have

$$\text{var}(\kappa_4) = \frac{24}{n}\sigma^8.$$

Example 9.5

To find the sampling variance of the coefficient of variation

$$V = \frac{100\sqrt{m_2}}{m_1'}$$

Taking logarithms and then differentials we have

$$\frac{dV}{V} = \frac{dm_2}{2m_2} - \frac{dm_1'}{m_1'}.$$

Whence, squaring and taking mean values,

$$\frac{\text{var } V}{V^2} = \frac{\text{var } m_2}{4m_2^2} - \frac{1}{m_2 m_1'}\text{cov}(m_2, m_1') + \frac{\text{var } m_1'}{m_1'^2} = \frac{1}{n}\left(\frac{\mu_4 - \mu_2^2}{4m_2^2} - \frac{\mu_3}{m_2 m_1'} + \frac{\mu_2}{m_1'^2}\right).$$

To our order of approximation we may write $\mu_r = m_r$ and find

$$\text{var } V = \frac{V^2}{n}\left(\frac{\mu_4 - \mu_2^2}{4\mu_2^2} - \frac{\mu_3}{\mu_2\mu_1'} + \frac{\mu_2}{\mu_1'^2}\right).$$

For a normal parent this gives ($\mu_4 = 3\mu_2^2$, $\mu_3 = 0$);

$$\text{var } V = \frac{V^2}{n}\left(\frac{1}{2} + \frac{\mu_2}{\mu_1'^2}\right) = \frac{V^2}{2n}\left(1 + \frac{2V^2}{10^4}\right)$$
$$= \frac{V^2}{2n} \text{ approximately.}$$

9.7. On the above principles the standard errors of the more usual functions of moments, such as the measures of skewness and kurtosis, have been worked out and tabulated (see *Tables for Statisticians and Biometricians* and the references at the end of the chapter).

In applications of results derived by the foregoing methods a few points are to be noted :

(*a*) The sampling variances are to be used only when the statistic under consideration is calculated from the moments. For instance, if the standard deviation of a normal curve is estimated by taking $\sqrt{\left(\frac{\pi}{2}\right)}$ times the mean deviation of the sample, instead of

the more usual root-mean-square, the formula var $(\sigma) = \frac{(\mu_4 - \mu_2^2)}{4n\mu_2}$ derivable from (9.9) is not applicable (see below, **9.11**).

(*b*) From (9.4) and (9.9) it will be seen that the sampling variance of a moment depends on the moment of twice the order, i.e. becomes very large for higher moments, even when n is large. This is the reason why such moments have very limited practical application.

(*c*) Some measures calculated from the moments tend to normality very slowly. $\sqrt{b_1}$ or b_1 (the sample values of $\sqrt{\beta_1}$ or β_1) are cases in point, and more refined methods which we discuss in Chapter 11 are preferable to the use of the standard error.

(*d*) The order of the approximation makes it necessary to exercise care in the neighbourhood of vanishing values of standard errors. For instance, if the coefficient of variation $V = 0$ in a sample, the formula of Example 9.5 would give var $V = 0$. But it does not, of course, follow that there is no variation at all in the population, though none exists in the sample and the presence of variation in the parent will be unlikely if the sample is at all large. When $V = 0$ the quantities neglected in our approximation giving var $V = \frac{V^2}{2n}$ become of some relative importance, though they are still small.

(*e*) It is interesting to compare the sampling fluctuations, as expressed in the sampling variance, with Sheppard's corrections to the moments. Writing temporarily s_1^2 for the uncorrected variance in the sample, s_2^2 for the corrected variance, we have

$$\frac{s_2^2}{s_1^2} = 1 - \frac{1}{12}\frac{h^2}{s_1^2},$$

where h is the interval width. For many practical cases, if d is the number of intervals, dh is about equal to $6s_1$, and thus

$$\frac{s_2^2}{s_1^2} = 1 - \frac{3}{d^2}$$

$$\frac{s_2}{s_1} = 1 - \frac{3}{2d^2} \text{ approximately.}$$

For a normal population we have

$$\sigma = \sqrt{\mu_2}$$

$$d\sigma = \frac{1}{2\sqrt{\mu}}\cdot d\mu_2$$

and hence

$$\text{var } \sigma = \frac{1}{4\mu_2} \text{ var } \mu_2$$

$$= \frac{\mu_4 - \mu_2^2}{4\mu_2 n}$$

$$= \frac{\mu_2}{2n} = \frac{\sigma^2}{2n}.$$

Thus if n is, say, 1000, the standard error of σ is about $0{\cdot}0224\,\sigma = 2{\cdot}24$ per cent. of σ. Sheppard's correction in a case where $d = 20$ is only 0·375 per cent. of s_1, i.e. only about a sixth of the standard error. It is as well to make the corrections, even when n is smaller than 1000, in order to avoid systematic error: but the correction should not be misinterpreted as implying a higher degree of reliability in the corrected value than actually exists.

Similar considerations apply *a fortiori* to the higher moments.

Standard Error of Bivariate Moments

9.8. Extensions of the above formulae to the bivariate case are made without difficulty, only slightly more complicated algebra being involved. The reader will be able to verify the following formulae for himself:

$$\text{var}\,(m'_{r,\,s}) = \frac{1}{n}(\mu'_{2r,\,2s} - \mu'^{2}_{r,\,s}) \quad . \tag{9.14}$$

$$\text{cov}\,(m'_{r,\,s}, m'_{u,\,v}) = \frac{1}{n}(\mu'_{r+u,\,s+v} - \mu'_{r,\,s}\,\mu'_{u,\,v}) \tag{9.15}$$

$$\begin{aligned}\text{var}\,(m_{r,\,s}) = \frac{1}{n}(&\mu_{2r,\,2s} - \mu^2_{r,\,s} + r^2\mu_{2,\,0}\,\mu^2_{r-1,\,s} + s^2\mu_{0,\,2}\,\mu^2_{r,\,s-1} \\ &+ 2rs\mu_{1,\,1}\,\mu_{r-1,\,s}\,\mu_{r,\,s-1} - 2r\,\mu_{r+1,\,s}\,\mu_{r-1,\,s} - 2s\mu_{r,\,s+1}\,\mu_{r,\,s-1})\end{aligned} \tag{9.16}$$

$$\begin{aligned}\text{cov}\,(m_{r,\,s}, m_{u,\,v}) = \tfrac{1}{2}(&\mu_{r+u,\,s+v} - \mu_{r,\,s}\,\mu_{u,\,v} + ru\mu_{20}\,\mu_{r-1,\,s}\,\mu_{u-1,\,v} \\ &+ sv\mu_{0,\,2}\,\mu_{r,\,s-1}\,\mu_{u,\,v-1} + rv\mu_{1,\,1}\,\mu_{r-1,\,s}\,\mu_{u,\,v-1}, \\ &+ su\mu_{1,\,1}\,\mu_{r,\,s-1}\,\mu_{u-1,\,v} - u\mu_{r+1,\,s}\,\mu_{u-1,\,v} \\ &- v\mu_{r,\,s+1}\,\mu_{u,\,v-1} - r\mu_{r-1,\,s}\,\mu_{u+1,\,v} - s\mu_{r,\,s+1}\,\mu_{u,\,v-1})\end{aligned} \tag{9.17}$$

Example 9.6

The coefficient of correlation is defined by

$$r = \frac{m_{1,\,1}}{\sqrt{(m_{20}m_{02})}}.$$

We have

$$\frac{dr}{r} = \frac{dm_{11}}{m_{11}} - \tfrac{1}{2}\frac{dm_{20}}{m_{20}} - \tfrac{1}{2}\frac{dm_{02}}{m_{02}}.$$

Thus

$$\frac{1}{r^2}\,\text{var}\,(r) = \frac{\text{var}\,(m_{11})}{m^2_{11}} + \tfrac{1}{4}\frac{\text{var}\,(m_{20})}{m^2_{20}} - \frac{\text{cov}\,(m_{11}, m_{20})}{m_{11}m_{20}} + \text{similar terms},$$

from which, substituting appropriate values from (9.16) and (9.17) and writing $\mu_{r,s}$ for $m_{r,s}$ in the result, we have

$$\text{var}\,(r) = \frac{\rho^2}{n}\left(\frac{\mu_{22}}{\mu^2_{11}} + \tfrac{1}{4}\frac{\mu_{40}}{\mu^2_{20}} + \tfrac{1}{4}\frac{\mu_{04}}{\mu^2_{02}} + \tfrac{1}{2}\frac{\mu_{22}}{\mu_{20}\mu_{02}} - \frac{\mu_{31}}{\mu_{11}\mu_{20}} - \frac{\mu_{13}}{\mu_{11}\mu_{02}}\right),$$

ρ being the same function of the μ's as r is of the m's. For the bivariate normal distribution the substitution of values of Example 3.15 gives

$$\text{var}\,(r) = \frac{1}{n}(1 - \rho^2)^2.$$

The use of the standard error to test the significance of the correlation coefficient is not, however, to be recommended.

Standard Errors of Quantiles

9.9. Among the various quantities measuring location and dispersion which we considered in Chapter 2 there was one group, namely the quantiles, which are not symmetric functions of the observations and whose sampling variances cannot accordingly be determined by the above methods. We proceed to consider them now.

Suppose the parent distribution is represented by $F(x) = f(x)\,dx$. The probability that, of a sample of n, $(l - 1)$ fall below a value x_1, one falls in the range $x_1 \pm \frac{1}{2}\,dx_1$ and the remaining $(n - l)$ fall above x_1 is proportional to

$$F(x_1)^{l-1}\,f(x_1)\,dx_1\{1 - F(x_1)\}^{n-l} = F_1^{\,l-1}(1 - F_1)^{n-l}\,dF_1, \quad . \quad . \quad . \tag{9.18}$$

where $F_1 = F(x_1)$. This expression is accordingly the distribution function of x_1, the member of the sample below which a proportion $\frac{l}{n}$ of the members fall, i.e. the lth quantile.

Put
$$l = nq$$
so that
$$n - l = n(1 - q) = np, \text{ say.}$$

The distribution (9.18) has a modal value given by differentiating the frequency function with respect to x_1, i.e. (taking logarithms first) by

$$(l-1)\frac{f_1}{F_1} + (n-l)\frac{f_1}{(1-F_1)} + \frac{f_1'}{f_1} \qquad (9.19)$$

this equation being satisfied by the modal value $\check{x}$. Now for large n, the factor $\frac{f_1'}{f_1}$ will in general be small compared with the other terms in (9.19), l and $n - l$ being large. We may therefore neglect it, and (9.19) becomes, to order n^{-1},

$$\frac{q}{F} + \frac{p}{1-F} = 0$$
or
$$F(\check{x}) = q.$$

This is in accordance with our general assumptions. To order n^{-1} the quantile of the sample is the quantile of the parent.

Now let us investigate the distribution (9.19) in the neighbourhood of the modal value. Put

$$F_1 = q + \xi.$$

(9.18) becomes (neglecting constants)

$$(q+\xi)^{nq}(p-\xi)^{np}.$$

Taking logarithms and expanding we have, except for constants,

$$nq \log\left(1 + \frac{\xi}{q}\right) + np \log\left(1 - \frac{\xi}{p}\right)$$
$$= nq\left(+\frac{\xi}{q} - \frac{1}{2}\frac{\xi^2}{q^2} \ldots\right) + np\left(-\frac{\xi}{p} - \frac{1}{2}\frac{\xi^2}{p^2} \ldots\right)$$
$$= -\frac{n\xi^2}{2pq} + \text{terms of order } \xi^3 \text{ and higher degree in } \xi.$$

Now for large samples ξ will be small compared with q, and we neglect the terms of higher order. Thus the distribution of ξ is

$$dF \propto \exp\left(\frac{-n\xi^2}{2pq}\right)d\xi$$

or, evaluating the necessary constant by integration,

$$dF = \frac{1}{\sqrt{2\pi}\sqrt{\left(\frac{pq}{n}\right)}} \exp\left(\frac{-n\xi^2}{2pq}\right)d\xi, \qquad (9.20)$$

showing that ξ is in the limit distributed normally with variance

$$\text{var}(\xi) = \frac{pq}{n}. \qquad (9.21)$$

This is the variance of ξ, which is a proportion. To find the variance of x_1 we note that $d\xi = dF_1 = f_1 dx_1$ and hence that

$$\text{var}\,(x_1) = \frac{pq}{nf_1^2}. \qquad (9.22)$$

In practice this formula is often applied to grouped frequency-distributions, and in such applications it is to be remembered that f_1, the ordinate of the parent, is to be taken as the frequency *per unit interval* at x_1, this being the best estimate of the ordinate.

Example 9.7

If x_1 is the median, $p = q = \frac{1}{2}$ and we have var (median) $= \dfrac{1}{4nf_1^2}$, where f_1 is the median ordinate. For instance, if the parent population is normal, the median ordinate is (from Appendix Table 1) $\dfrac{1}{\sigma}\,0{\cdot}39894$, σ^2 being the variance of the parent. Hence the standard error of the median is

$$\frac{\sigma}{\sqrt{n}}\cdot\frac{1}{2 \times 0{\cdot}39894}$$

$$= 1{\cdot}2533\frac{\sigma}{\sqrt{n}}.$$

The standard error of the mean in samples of n from a normal population is $\dfrac{\sigma}{\sqrt{n}}$, which is thus considerably smaller than the standard error of the median.

9.10. To find the covariance of two quantiles we generalise equation (9.18). If we have a random sample of n individuals the probability that $(l - 1)$ lie below x_1, one lies at $x_1 \pm \frac{1}{2}dx_1$, $(n - l - m)$ lie between x_1 and x_2, one at $x_2 \pm \frac{1}{2}dx_2$, and the remaining $(m - l)$ above x_2 is

$$dF \propto F_1^{l-1}(F_2 - F_1)^{n-l-m}(1 - F_2)^{m-1}dF_1\,dF_2. \qquad (9.23)$$

where $F_1 = F(x_1)$, $F_2 = F(x_2)$.
We put

$$l = q_1 n$$
$$m = p_2 n$$

and find, for the equations giving the modal values corresponding to (9.19),

$$\frac{q_1}{p_1} - \frac{(q_2 - q_1)}{F_2 - F_1} = 0$$

$$\frac{q_2 - q_1}{F_2 - F_1} \quad \frac{p_2}{1 - F_2} = 0$$

giving, for the limiting modal values,

$$F(\check{x}_1) = q_1$$
$$F(\check{x}_2) = q_2 \qquad (9.24)$$

The conditions as to the relative smallness of $\dfrac{f'(\check{x})}{f(\check{x})}$ are satisfied in any ordinary case. Now put

$$F_1 = q_1 + \xi_1$$
$$F_2 = q_2 + \xi_2.$$

The joint distribution of ξ_1 and ξ_2 then becomes

$$dF \propto (q_1 + \xi_1)^{q_1 n}(q_2 - q_1 + \xi_2 - \xi_1)^{(q_2 - q_1)n}(p_2 - \xi_2)^{p_2 n}\, d\xi_1\, d\xi_2.$$

On proceeding as in the previous section, taking logarithms, expanding and neglecting terms in ξ^3 and higher, we find ultimately

$$dF \propto \exp\left\{-\frac{n}{2(q_2 - q_1)}\left(\frac{q_2}{q_1}\xi_1^2 - 2\xi_1\xi_2 + \frac{p_1\xi_2^2}{p_2}\right)\right\} d\xi_1 d\xi_2. \quad . \quad . \quad (9.25)$$

Thus the joint distribution of ξ_1 and ξ_2 tends to the bivariate normal form, and on comparing (9.25) with the canonical form (Example 3.15) we see that

$$\frac{1}{(1-\rho^2)\operatorname{var}(\xi_1)} = \frac{nq_2}{(q_2 - q_1)q_1}$$

$$\frac{1}{(1-\rho^2)\operatorname{var}(\xi_2)} = \frac{np_1}{(q_2 - q_1)p_2}$$

$$\frac{}{(1-\rho^2)\operatorname{cov}(\xi_1\xi_2)} = \frac{}{(1-\rho^2)\sqrt{\{\operatorname{var}(\xi_1)\operatorname{var}(\xi_2)\}}} = \frac{}{(q_2 - q_1)},$$

whence it is easy to find

$$\operatorname{var}(\xi_1) = \frac{p_1 q_1}{n}$$

$$\operatorname{var}(\xi_2) = \frac{p_2 q_2}{n}$$

$$\operatorname{cov}(\xi_1, \xi_2) = \frac{p_2 q_1}{n}. \quad . \quad (9.26)$$

The asymmetry of the result for the covariance is due to the fact that p_2 relates necessarily to the *upper* quantile. For the corresponding expression in x_1 and x_2 we have

$$\operatorname{cov}(x_1, x_2) = \frac{p_2 q_1}{n f_2 f_1}. \quad (9.27)$$

With equations (9.26) and (9.27) we can find expressions for the variances of the quantile range and similar statistics.

Example 9.8

The variance of the difference δ of two quantiles at x_1 and x_2 is given by

$$d\delta = dx_1 - dx_2,$$

$$\operatorname{var}(\delta) = \operatorname{var}(x_1) + \operatorname{var}(x_2) - 2\operatorname{cov}(x_1, x_2)$$

$$= \frac{1}{n}\left\{\frac{p_1 q_1}{f_1^2} + \frac{p_2 q_2}{f_2^2} - \frac{2p_2 q_1}{f_1 f_2}\right\}.$$

When the quantiles are the two quartiles, $p_2 = q_1 = \frac{1}{4}$, $p_1 = q_2 = \frac{3}{4}$, and for the variance of the *semi*-interquartile range we have

$$\operatorname{var}(\text{s.i.q.}) = \frac{1}{64n}\left(\frac{3}{f_1^2} + \frac{3}{f_2^2} - \frac{2}{f_1 f_2}\right),$$

where f_1, f_2, are the frequencies per unit interval at the two quartiles, f_2 relating to the upper quartile. As f_1, f_2 have to be estimated from the sample, we may also write

$$\operatorname{var}(\text{s.i.q.}) = \frac{\sigma^2}{64n}\left(\frac{3}{g_1^2} + \frac{3}{g_2^2} - \frac{2}{g_1 g_2}\right),$$

where g_1, g_2 are the actual sample frequencies at the quartiles and σ^2 is the sample variance.

For instance, if the parent distribution is normal, $g_1 = g_2$ and we find

$$\text{var (s.i.q.)} = \frac{\sigma^2}{16ng_1^2}.$$

From the tables the deviate corresponding to the quartile is 0·6745 and the ordinate at this point, $g_1 = 0{\cdot}3178$, so that the standard error of the semi-interquartile range is

$$\frac{}{\sqrt{n}\,.\,4 \times 0{\cdot}3178}$$

$$= 0{\cdot}7867\frac{\sigma}{\sqrt{n}}.$$

9.11. In amplification of the point mentioned in **9.7** (*a*) it is worth while stressing again the fact that a standard error is related to the way in which a parameter is estimated. For instance, the standard deviation of a normal curve can be estimated from a sample in several ways: from the second moment; by taking $\sqrt{\left(\frac{\pi}{2}\right)}$ times the mean deviation; by taking $\frac{1}{0.6745}$ times the semi-interquartile range; and so on. Each method will have its appropriate standard error, that for the first, for example, being $\frac{\sigma}{\sqrt{(2n)}}$, and that for the third $\frac{1{\cdot}6495\sigma}{\sqrt{(2n)}}$. At a later stage considerations such as this will lead us to the inquiry, what is the estimate, if any, with the *minimum* sampling variance? For present purposes it is enough to note the importance of not using a quoted formula without reference to the method of estimation of the parameter concerned.

9.12. The methods we have developed provide the standard errors for large samples of most of the measures of location and dispersion and the measures introduced in Chapters 2 and 3. There remain a few on which we have not yet touched, viz. the mean deviation, Gini's coefficient of mean difference, and the range. We consider them briefly in turn.

Standard Error of the Mean Deviation

9.13. The mean deviation, as was pointed out in Chapter 2, is relatively speaking a complicated function, and the mathematical difficulties attendant on absolute values are well illustrated in discussions of its sampling variance. In fact, no general discussion of the sampling distribution appears to have been undertaken. The following exact value of the sampling variance in samples from a normal population was discovered by Helmert in 1876 and rediscovered by Fisher in 1920.

$$\text{var (m.d.)} = \frac{2}{\pi}\frac{(n-1)}{n^2}\sigma^2\left(\frac{\pi}{2} + \sqrt{\{n(n-2)\}} - n + \sin^{-1}\frac{1}{n-1}\right)$$

$$\frac{\sigma^2}{n}\left(1 - \frac{2}{\pi}\right) \text{ for large } n. \quad \ldots \ldots \quad (9.28)$$

The proof follows the general methods described in the next chapter. It is quoted here for the sake of completeness.

Standard Error of the Mean Difference

9.14. Nair (1936) has given a general expression for the standard error of Gini's mean difference without repetition. In the manner of **2.24** it is easy to see that the coefficient may be written

$$\Delta_1 = \frac{2}{n(n-1)}\{2U - (n+1)V\} \quad . \quad . \quad . \quad . \quad (9.29)$$

where

$$U = \sum_{j=1}^{n}(jx_j)$$

$$V = \sum_{j=1}^{n} x_j,$$

and we write n in preference to N for the number of observations, since we are dealing with a sample.

In our usual notation, the probability that the jth observation in order of magnitude in a series of n observations has value in the range $x \pm \frac{1}{2}dx$ is

$$dF = \frac{n!}{(j-1)!(n-j)!} F_1^{j-1}(1 - F_1)^{n-j}\, dF_1.$$

Hence the mean value of U is given by

$$E(U) = \int_{-\infty}^{\infty} \sum_{j=1}^{n}\left\{jx\frac{n!}{(j-1)!(n-j)!}F^{j-1}(1-F)^{n-j}\right\}dF$$

$$= n\int_{-\infty}^{\infty} xdF\left\{\Sigma\frac{(n-1)!j}{(j-1)!(n-j)!}F^{j-1}(1-F)^{n-j}\right\}$$

$$= n\int_{-\infty}^{\infty} xdF\left\{\Sigma\frac{(n-1)!}{\overline{j-1}!(n-j)!}F^{j-1}(1-F)^{n-j}\right.$$

$$\left. + (n-1)F\,\Sigma\frac{(n-2)!}{\overline{j-2}!\,\overline{n-j}!}F^{j-2}(1-F)^{n-j}\right\}$$

$$= n\int_{-\infty}^{\infty} xdF\{1 + (n-1)F\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.30)$$

Similarly

$$E(V) = n\int xdF. \quad . \quad . \quad (9.31)$$

Thus

$$E(\Delta_1) = \frac{\quad}{n(n-1)}\{2E(U) - (n+1)E(V)\}$$

$$= 2\int x(2F - 1)\, dF.$$

In the same way (but we omit the details) Nair finds

$$E(\Delta_1^2) = \frac{\quad}{n(n-1)}\{I_1 + 2(n-2)I_2\} \quad . \quad . \quad (9.32)$$

where

$$I_1 = \int_{-\infty}^{\infty} x^2\{(n-1) - 4(n-2)F + 4(n-2)F^2\}dF$$

$$I_2 = \int_{-\infty}^{\infty} x_2\, dF_2 \int_{-\infty}^{x_2} x_1\, dF_1\{(n-3) - 2(n-5)F_1 - 2(n-1)F_2 + 4(n-3)F_1F_2\}$$

and finally

$$\text{var}\,(\Delta_1) = E(\Delta_1^2) - \{E(\Delta_1)\}^2 \quad . \quad . \quad . \qquad (9.33)$$

For three particular cases these integrals are worked out, giving:

Normal Parent:

$$dF = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} dx, \qquad -\infty \leqslant x \leqslant \infty$$

$$E(\Delta_1) \quad \frac{2\sigma}{\sqrt{\pi}}$$

$$\text{var}\,(\Delta) = \frac{4\sigma^2}{n(n-1)}\left\{\frac{n+1}{3} + \frac{2(n-2)\sqrt{3}}{\pi} - \frac{2(2n-3)}{\pi}\right\} \quad . \quad . \qquad (9.34)$$

$$\sim \frac{\sigma^2}{n}(0{\cdot}8068)^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9.35)$$

Exponential Parent:

$$dF = \frac{1}{\sigma} e^{\frac{-x}{\sigma}} dx, \qquad 0 \leqslant x \leqslant \infty$$

$$E(\Delta_1) = \sigma \quad . \quad . \quad . \quad . \quad . \qquad (9.36)$$

$$\text{var}\,(\Delta_1) = \sigma^2\frac{2(2n-1)}{3n(n-1)} \quad . \quad . \quad . \quad . \qquad (9.37)$$

$$\sim \frac{4}{3n}\sigma^2. \quad . \quad . \quad . \quad . \quad . \qquad (9.38)$$

Rectangular Parent:

$$dF = \frac{1}{k}dx \qquad 0 \leqslant x \leqslant k$$

$$E(\Delta_1) = \tfrac{1}{3}k \quad . \quad . \quad . \qquad (9.39)$$

$$\text{var}\,(\Delta_1) = \frac{k^2}{9}\cdot\frac{n+3}{5n(n-1)} \quad . \qquad (9.40)$$

$$\sim \frac{k^2}{45n}. \quad . \quad . \qquad (9.41)$$

9.15. We now turn to consider some statistics which are peculiar in more ways than one—the extreme values of a sample (or, more generally, the mth value from the top or the bottom of a sample) and the range. One of the unusual features of the distribution of mth values is that as n increases it *diverges* more and more from normality; and it seems doubtful whether the distribution of range tends to any limit at all—certainly it does not tend to normality in all cases.

A further difference between the quantities we are now considering and the others we have

already discussed is that mth values and range in the sample are not used to estimate mth values and range in the population. In fact most of the results we shall obtain relate to parents which have an infinite range. What, then, is the use of these statistics? The answer is that they may provide an estimate of parent parameters which do exist. For instance, an estimate of the variance of a normal population is given by dividing the sample range w by a constant d_n depending on the number in the sample. This estimate, though not so accurate as some (in the sense that its sampling variance is not so small), is extremely easy to calculate and is often useful. We wish, therefore, to know its sampling variance, that is to say the sampling variance of the range.

*Distribution of m*th *Values*

9.16. We consider first of all the distribution of mth values from the top, that for mth values from the bottom being similar. In particular, m may be unity, in which case we get the greatest member of a sample.

Quantiles are special cases of this class of statistic, the ratio m/n remaining finite as n tends to infinity. In the case we now discuss m remains finite, so that the ratio m/n tends to zero.

The distribution of mth values from the top is, as in equation (9.18),

$$dF \propto F_1^{n-m}(1 - F_1)^{m-1}\, dF_1. \qquad . \quad . \quad . \quad . \; (9.42)$$

When the form of F_1 is known, this equation is sometimes capable of exact solution, as in the following example.

Example 9.9

Consider the rectangular distribution $dF = dx$, $0 \leqslant x \leqslant 1$. Here $F(x) = x$ and the distribution of the mth value from the top is

$$dF \propto x^{n-m}(1 - x)^{m-1}\, dx,$$

the Pearson Type I curve. We have, for the first moment,

$$\mu_1 = \frac{n - m + 1}{n + 1} = 1 - \frac{m}{n + 1}$$

and for the variance

$$\mu_2 = \frac{(n - m + 1)(2n - m)}{(n + 1)(n + 2)},$$

but this sampling variance cannot be used in the ordinary way if m is finite, for the curve does not tend to normality. However, we may easily obtain exact values for the probabilities associated with the mth values, from the integrals of the Type I curve. In fact, the probability that a given value will not be attained or exceeded is, in the usual notation, $I_x(n - m + 1, m)$.

9.17. From this point our discussion of the limiting form of (9.42) as n tends to infinity is confined to the case wherein the parent population is a continuous frequency-distribution of unlimited range of the exponential type, i.e. such that it tends to zero with large x as fast as or faster than $dF = e^{-|x|}\, dx$, and that $\dfrac{d}{dx}\log f(x)$ exceeds some fixed number as x tends to infinity. The normal curve obeys this criterion, which implies, among other things, that all moments exist.

For the mode of (9.42) we have, as in (9.19),

$$(n-m)\frac{f_1}{F_1}-\frac{(m-1)f_1}{1-F_1}+\frac{f_1'}{f_1}=0.$$

For large n and finite m the mode x_1 will be a large value, and both f_1' and $1-F_1$ tend to zero. Accordingly we may put

$$-\frac{f_1'}{f_1}=\frac{\dfrac{d}{dx}f_1}{\dfrac{d}{dx}(1-F_1)}\sim\frac{f_1}{1-F_1}$$

in accordance with the rule known as L'Hôpital's. Hence

$$(n-m)\frac{f_1}{F_1}-m\frac{f_1}{1-F_1}=0$$

$$F_1(\check{x})=1-\frac{m}{n}.$$

Now expand F_1 in the neighbourhood of $\check{x}$ by Taylor's theorem. We get

$$F(x)=F(\check{x})+\frac{x-\check{x}}{1!}f(\check{x})+\frac{(x-\check{x})^2}{2!}f'(\check{x})+\dots$$

$$=1-\frac{m}{n}+\frac{m}{n}(x-\check{x})\frac{n}{m}f(\check{x})-\frac{m}{n}\frac{(x-\check{x})^2}{2!}\frac{n^2}{m^2}\{f(\check{x})\}^2+\dots$$

(the last term in virtue of $f_1'\sim\dfrac{-f_1^2}{1-F_1}=\dfrac{-f_1^2}{\frac{n}{m}}$ in the neighbourhood of the mode)

$$=1-\frac{m}{n}\exp\left\{-(x-\check{x})\frac{n}{m}f(\check{x})\right\}\text{approximately} \quad . \quad . \qquad (9.43)$$

$$=1-\frac{m}{n}e^{-y_m},\ \text{say.} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9.44)$$

The distribution of the mth value from the top may be written, from (9.42),

$$dF_m\propto\left(\frac{1}{F}-1\right)^{m-1}d(F^n).$$

To our approximation, from (9.44), since $\dfrac{m}{n}$ is small,

$$\left(\frac{1}{F}-1\right)^{m-1}=\left(\frac{1}{1-\frac{m}{n}e^{-y_m}}-1\right)^{m-1}$$

$$\sim\left(\frac{m}{n}\right)^{m-1}e^{-(m-1)y_m}$$

and $$d(F^n)=n\left(1-\frac{m}{n}e^{-y_m}\right)^{n-1}\frac{m}{n}e^{-y_m}dy_m$$

Thus $$dF_m\propto\exp(-my_m-me^{-y_m})\,dy_n$$

and, on evaluating the constants by integration, we get

$$dF_m=\frac{m^m}{(m-1)!}\exp(-my_m-me^{-y_m})dy_n \qquad . \ (9.45)$$

The new variable y_m is defined in terms of $x - \check{x}$ by

$$y_m = (x - \check{x})\cdot\frac{nf(\check{x})}{m}$$

In a similar way, for the mth value from the bottom we find

$$dF = \frac{m^m}{(m-1)!}\exp\,(m_m y - me^{_m y})\,dy, \tag{9.46}$$

$_m y$ being written for the variable defined by

$$F = \frac{m}{n}e^{_m y}.$$

In particular, for the extremes $(m = l)$ we have

$$dF = \exp\,(-\,y - e^{-y})\,dy \text{ (top value)} \tag{9.47}$$

$$dF = \exp\,(y - e^{y})\,dy \text{ (bottom value)} \tag{9.48}$$

9.18. These unusual limiting forms, which are due to Gumbel (1934), the extreme cases being due to Fisher and Tippett (1928), are very far from normal for moderate or low values of m. For the moments of (9.45) we have (omitting the suffix of y for convenience)

$$\mu_1' = \frac{m^m}{(m-1)!}\int_{-\infty}^{\infty} e^{-my-me^{-y}}\,y\,dy.$$

Put $\quad e^{-y} = \frac{t}{m}.$ We get

$$\begin{aligned}\mu_1' &= \frac{1}{(m-1)!}\int_0^{\infty}(\log t - \log m)t^{m-1}\,e^{-t}\,dt\\ &= -\log m + \frac{d}{dm}\{\log \Gamma(m)\}\\ &= -\log m + \gamma - \sum_{r=1}^{m-1}\left(\frac{1}{r}\right)\end{aligned} \tag{9.49}$$

where γ is Euler's constant. For the rth moment about the mean we have

$$\begin{aligned}\mu_r &= \frac{(-1)^r}{(m-1)!}\int_0^{\infty}(\log t - \log m + \mu_1')t^{m-1}\,e^{-t}\,dt\\ &= \frac{(-1)^r}{\Gamma(m)}\left[\frac{d^r}{dq^r}e^{q(\mu'_1 - \log m)}\,\Gamma(m+q)\right]_{q=}\end{aligned} \tag{9.50}$$

These formulae have been worked out further by Gumbel, from whose numerical results the following are chosen :—

m	Mean	β_1	β_2-3
1	0·577	1·139	2·400
3	0·176	0·621	0·763
5	0·103	0·468	0·437
10	0·051	0·324	0·212

These figures, which relate to the distribution from the top, show clearly that the limiting distribution is far from normal. The distribution from the bottom is similar, odd moments including the mean having the same magnitude but opposite sign, even moments being the same.

Moreover, the limiting forms (9.45) and (9.46) are reached extremely slowly. Fisher and Tippett (1928) have shown in the case $m = 1$ that they do not provide a very satisfactory approximation for values of n less than 10^{12}. For practical purposes, therefore, there is still no adequate general approximate form for the distribution of mth values.

9.19. The case $m = 1$, corresponding to the extremes of the sample, has, however, been studied in more detail. In this case equation (9.42) becomes

$$dF = \frac{d}{dx_1} F_1{}^n \, dx_1.$$

By using the published tables of the normal integral F_1, Tippett (1925) has evaluated F for values of n up to 1000, and given diagrams yielding the variances, and β_1 and β_2, which are reproduced in *Tables for Statisticians and Biometricians*, Part II. The following values are quoted from his results :—

n	Mean	Standard Deviation	β_1	β_2
2	0·564	0·826	0·019	3·062
5	1·163	0·669	0·092	3·202
10	1·539	0·587	0·168	3·331
100	2·508	0·429	0·429	3·765
500	3·037	0·370	0·570	4·003
1000	3·241	0·351	0·618	4·088

The values of β_1 and β_2 illustrate the point that as n increases, the distribution of the extreme value diverges more and more from the normal form.

The limiting values as $n \rightarrow \infty$ can be derived by the use of characteristic functions. In fact, we have for the distribution of the top value,

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \exp\,(-x - e^{-x})\,dx,$$

which, on substituting $e^{-x} = \xi$, gives

$$\phi(t) = \int^{\infty} \xi^{-it} e^{-\xi} \, d\xi$$
$$= \Gamma(1 - it).$$

Hence

$$\kappa_1(it) + \frac{\kappa_2(it)^2}{2!} \quad . \; . = \log \phi(t) = \log \Gamma(1 - it)$$
$$= \gamma(it) \quad \frac{S_2}{2}(it)^2 \quad \frac{S_3}{3}(it)^3 \qquad \text{etc.*}$$

Thus

$$\kappa_1 = \mu_1$$
$$\kappa_2 = \mu_2 = S_2 = \frac{\pi^2}{6} = 1\cdot644934$$
$$\mu_3 = 2S_3 = 2\cdot404114$$
$$\mu_4 - 3\mu_2^2 = 6S_4 = \frac{\pi^4}{15} = 6\cdot493939$$

whence

$$\beta_1 = 1\cdot299$$
$$\beta_2 = 5\cdot4.$$

* Cf. Edwards, *Integral Calculus*, vol. 2, article 916. S_r here is $\sum_{n=1}^{\infty} \frac{1}{n^r}$.

These are evidently far from equal to the values for $n = 1000$ given above. Clearly the limiting form is an inadequate approximation for values of n much higher than 1000.

9.20. The problem of bridging the gap between Tippett's values and the limiting form has been considered by Fisher and Tippett (1928), and the argument which they employ is interesting. Concentrating for a moment on the upper value, we note that the upper member of a sample of kn members is the upper member of a sample of k of upper members of samples of n. Both distributions will tend to the same limiting form, if it exists; and consequently the limiting value must be such that the extreme member of a sample of n from it must itself have that distribution. That is to say, if F is the probability of an observation being less than x,

$$F^n(x) = F(a_n x + b_n), \quad . \quad . \quad . \quad . \quad . \quad (9.51)$$

where a_n and b_n are functions of n.

It may be shown from this equation that F must be one of three forms:—

$$dF = \exp(-x - e^{-x})dx \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.52)$$

$$dF = \frac{A}{x^{A+1}} \exp(-x^{-A})\, dx \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.53)$$

$$dF = A(-x)^{A-1} \exp\{-(-x)^A\}dx \quad . \quad . \quad . \quad . \quad (9.54)$$

The first we have already reached. The second and third arise if the original distribution, instead of tending to infinity exponentially, tends less rapidly such that

$$\lim_{x \to 0} (1 - F)x^A \text{ exists and is not zero.}$$

The distribution (9.54) itself has (9.52) as a limiting form as A tends to infinity. It has therefore been proposed as a "penultimate" form, to bridge the gap between $n = 1000$ and $n = 10^{12}$, which is apparently the first point at which the ultimate form provides a reasonable approximation. For the penultimate form we have

$$\mu'_r = \int_{-\infty}^{\infty} A(-x)^{A-1} x^r e^{-(-x)^A}\, dx$$

and on putting

$$-x = t^{\frac{1}{A}}$$

$$\mu'_r = \int_0^{\infty} (-1)^r t^{\frac{r}{A}} e^{-t} dt$$

$$= (-1)^r \Gamma\left(1 + \frac{r}{A}\right).$$

The following values illustrate the relationship between the known form ($n = 500$, 1000) and the penultimate form:

$\frac{1}{A}$	n	Standard Deviation		β_1		β_2	
		Penultimate	Actual	Penultimate	Actual	Penultimate	Actual
0·0768	1000	0·3433	0·3514	0·548	0·618	3·852	4·088
0·0845	500	0·3604	0·3704	0·498	0·570	3·751	4·003

Distribution of Range

9.21. The range is the difference of the highest and the lowest value of a sample, and the simultaneous distribution of top and bottom values is, from (9.23),

$$\begin{aligned} dF &\propto (F_1 - F_2)^{n-2}\, dF_1\, dF_2 \\ &= n(n-1)(F_1 - F_2)^{n-2}\, dF_1\, dF_2. \qquad (9.55) \end{aligned}$$

The distribution function of the range w is then given by integrating this distribution over values of F_1 and F_2 such that $x_2 - x_1 \leqslant w$. So far as I am aware, it is not known whether limiting forms of this distribution exist or what they are. It is, however, evident that for large n the range is also large, and it seems doubtful whether the difference of two variates which (for an unlimited curve) tend to $+\infty$ and $-\infty$ respectively has any general limiting form. In any case one would suspect that the limiting form is reached slowly.

For particular cases equation (9.55) is soluble explicitly. The normal case has been fairly completely studied by Tippett (1925) and E. S. Pearson (1926 and 1932). Tippett found the first four moments of the distribution of the range, tabulated the mean values for values of n up to 1000 and gave a diagram for determining standard errors. (These tables and diagram are reproduced in *Tables for Statisticians and Biometricians*, Part II.) Briefly, his approach is as follows :—

From (9.55) we have, for the mean range $E(w)$,

$$E(w) = n(n-1)\int_{-\infty}^{\infty} dF_1 \int_{-\infty}^{x_1} (F_1 - F_2)^{n-2}(x_1 - x_2)\, dF_2. \qquad (9.56)$$

On expanding $(F_1 - F_2)^{n-2}$ we get terms under the second integral sign like

$$\begin{aligned} \int_{-\infty}^{x_1} F_2^{S}(x_1 - x_2)\, dF_2 &= \left[\frac{(x_1 - x_2)F_2^{S+1}}{S+1}\right]_{-\infty}^{x_1} + \frac{1}{S+1}\int_{-\infty}^{x_1} F_2^{S-1}\, dx_2 \\ &= \frac{1}{S+1}\int_{-\infty}^{x_1} F_2^{S+1}\, dx_2 = \frac{U^{S+1}}{S+1} \text{ say.} \end{aligned}$$

Then
$$E(w) = n!\sum_{S=0}^{n-2}\frac{(-1)^S}{(S+1)!(n-2-S!)}\int_{-\infty}^{\infty} F_1^{n-2-S}\, U_{S-1}\, dF_1.$$

But $$\int_{-\infty}^{\infty} F_1^{n-2-S}\, U_{S+1} dF_1 = \left[\frac{U_{S+1}F_1^{n-1-S}}{n-1-S}\right]_{-\infty}^{\infty} - \frac{1}{n-1-S}\int_{-\infty}^{\infty} F_1^{n-1-S}\, F_1^{S+1}\, dF_1$$

Hence

$$\begin{aligned} E(w) &= n!\sum_{S=0}^{n-2}\frac{(-1)^S}{(S+1)!(n-1-S!)}\int_{-\infty}^{\infty} (1 - F_1^{n-1-S})F_1^{S+1}\, dF_1 \\ &= \int_{-\infty}^{\infty} \{1 - (1 - F_1)^n - F_1^n\}\, dx_1. \qquad (9.57) \end{aligned}$$

In a similar way it is found that

$$\begin{aligned} \text{var }(w) = 2\int_{-\infty}^{\infty}\int_{-\infty}^{x_1} &\{1 - F_1^n - (1 - F_2)^n - (F_1 - F_2)^n\}\, dx_1\, dx_2 \\ &- \{E(w)\}^2. \qquad (9.58) \end{aligned}$$

This equation was used by Tippett to obtain values of the variance for n up to 1000. The following values illustrate the general behaviour of the distribution :—

n	Standard Deviation	β_1 (approximate)	β_2 (approximate)
2	0·853	0·99	3·87
10	0·797	0·16	3·15
100	0·605	0·21	3·38
500	0·524	0·29	3·50
000	0·497	0·31	3·54

Again it would appear that as n increases, the distribution of range diverges more and more from the normal form.

The distribution function of the range in normal samples has recently been tabulated by E. S. Pearson and Hartley (1942).

List of Standard Errors of Commonly Occurring Statistics

9.22. In view of the general utility of the standard error it may be convenient to bring together at this point for reference a number of sampling variances and other results. Some of these have already been obtained in this chapter ; others are direct consequences of the formulae or methods developed ; and some will be proved later in the book.

Mean. $\text{var}\,(m_1') = \frac{\mu_2}{n} = \frac{\sigma^2}{n}$ where σ is the standard deviation of the parent. This is true in particular for a normal parent. The mean is always estimated from the mean of the sample.

Variance. $\text{var}\,(m_2) = \frac{(\mu_4 - \mu_2^2)}{n}$. For the normal parent $\text{var}\,(m_2) = \frac{2\sigma^4}{n}$. Tables are given for this case in T.S.B. I *. These results are appropriate to the case where the variance is estimated from the sample variance. For numerical results for other cases see Davies and E. S. Pearson (1934).

Standard Deviation. $\text{var}\,(s) = \frac{(\mu_4 - \mu_2^2)}{4n\mu_2}$. For normal parent $\text{var}\,(s) = \frac{\sigma^2}{2n}$. These are the values for estimates from the square root of the sample variance. See previous note on variance.

Third Moment about the Mean. $\text{var}\,(m_3) = \frac{(\mu_6 - \mu_3^2 - 6\mu_4\mu_2 + 9\mu_2^3)}{n}$. For normal parent $\text{var}\,(m_3) = \frac{6\sigma^6}{n}$. The third and higher moments are always estimated from the moments of the sample.

Fourth Moment about the Mean. $\text{var}\,(m_4) = \frac{(\mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_2\mu}{n}$ For normal parent $\text{var}\,(m_4) = \frac{96\sigma^8}{n}$

Coefficient of Variation. $\text{var}\,(V) = \frac{V^2}{n}\left(\frac{\mu_4 - \mu_2^2}{4\mu_2^2} + \frac{\mu_2}{\mu_1'^2} - \frac{\mu_3}{\mu_2\mu_1'}\right)$. For normal parent $\text{var}\,(V) = \frac{V^2}{2n}$ approximately. Tables given in T.S.B. I.

* An abbreviation for *Tables for Statisticians and Biometricians*, Part I.

β_1. $\text{var}\,(\beta_1) = \dfrac{\beta_1(4\beta_4 - 24\beta_2 + 36 + 9\beta_1\beta_2 - 12\beta_3 + 35\beta_1)}{n}$. For normal parent var $(\beta_1) = \dfrac{6\beta_1}{n}$. Tables given in T.S.B. I. The distribution is fairly skew for moderately large n and the methods of Chapter 11 provide better tests of β_1 as a measure of departure from normality. See **11.23.** (The β's are defined in equation (3.65).)

β_2. $\text{var}\,(\beta_2) = \dfrac{(\beta_6 - 4\beta_2\beta_4 + 4\beta_2^3 - \beta_2^2 + 16\beta_2\beta_1 - 8\beta_3 + 16\beta_1)}{n}$. Tables given in T.S.B. I.

Pearson Measure of Skewness (Equation (3.64)). Tables given in T.S.B. I. Probably skew for moderate n. See note on β_1.

Pearson Mode. Formulae and tables given in Yasukawa (1926), the results of course being only applicable to modes calculated from the Pearson formula (equation (3.62)). Distribution may be skew for moderate n.

Coefficient of Contingency. See **13.14.**

Coefficient of Association. See **13.8.**

Tetrachoric r. See **14.28.**

Mean Deviation. General formulae not known. See **9.13.** For normal parent $\text{var (m.d.)} = \dfrac{\sigma^2}{n}\left(1 - \dfrac{2}{\pi}\right)$.

Gini's Mean Difference. See **9.14.** For normal case $\text{var}\,(\varDelta_1) = \dfrac{(0{\cdot}8068)^2\sigma^2}{n}$.

Median. $\text{var}\,(m_e) = \dfrac{\sigma^2}{4ny_0^2}$ where y_0 is the median ordinate of the sample. For normal parent $\text{var}\,(m_e) = \dfrac{(1{\cdot}2533)^2\sigma^2}{n}$. For small samples from normal population, tables and formulae given in Hojo (1931). Results to higher order in n given by K. Pearson (1931).

Quartiles. $\text{var}\,(Q) = \dfrac{3\sigma^2}{(4ny^2)}$ where y is ordinate at the quartile concerned. For normal parent, $\text{var}\,(Q) = \dfrac{(1{\cdot}3626)^2\sigma^2)}{n}$. Results for small samples from normal population given in Hojo (1931).

Semi-interquartile range. $\text{var (s.i.q.)} = \dfrac{\sigma^2}{4n}\left(\dfrac{3}{16y_1^2} + \dfrac{3}{16y_2^2} - \dfrac{1}{8y_1y_2}\right)$ where y_1, y_2, are the quartile ordinates. For normal parent $\text{var (s.i.q.)} = \dfrac{(0{\cdot}7867)^2\sigma^2}{n}$

Deciles. For the normal parent, variances are

$$\text{for deciles } 4,\ 6 \quad \frac{(1{\cdot}2680)^2\sigma^2}{n}$$

$$3,\ 7 \quad \frac{(1{\cdot}3180)^2\sigma^2}{n}$$

$$2,\ 8 \quad \frac{(1{\cdot}4288)^2\sigma^2}{n}$$

$$1,\ 9 \quad \frac{(1{\cdot}7094)^2\sigma^2}{n}$$

Range. See **9.21**.

Correlation Coefficient. See **14.10**. For normal case var $(r) = \dfrac{(1-\rho^2)^2}{n}$. But it is better to use Fisher's transformation (**14.18**) or the Tables by David (1938).

Coefficient of Regression. See **14.10** and **14.11**. For normal case var $(b_1) = \dfrac{\sigma_1^2(1-\rho^2)}{\sigma_2^2 n}$.

Standard Errors of Sums and Differences

9.23. Suppose we have two variables x_1, x_2, which may or may not be independent. We have, if z is their sum,

$$E(z) = E(x_1) + E(x_2),$$

or the mean of z is the sum of the means of x_1 and x_2. If then we measure x_1 and x_2 about their respective means, the mean of z is zero and thus

$$\begin{aligned}\text{var } z = E(z^2) &= E(x_1 + x_2)^2 \\ &= E(x_1^2) + 2E(x_1x_2) + E(x_2^2) \\ &= \text{var } x_1 + 2 \text{ cov } (x_1, x_2) + \text{var } x_2. \qquad (9.59)\end{aligned}$$

Similarly for the difference of two variables we have

$$\text{var } z = \text{var } x_1 - 2 \text{ cov } (x_1, x_2) + \text{var } x_2 . \qquad (9.60)$$

In particular, if x_1 and x_2 are independent their covariance vanishes, for it becomes the product of the two means, each of which is zero. In this important case we have, for the sum,

$$\text{var } (x_1 + x_2) = \text{var } x_1 + \text{var } x_2 \qquad (9.61)$$

and for the difference

$$\text{var } (x_1 - x_2) = \text{var } x_1 + \text{var } x_2. \qquad (9.62)$$

These results are of fundamental importance : the variance of the sum or difference of two independent random variables is the sum of their variances. Generally if

$$z = a_1x_1 + a_2x_2 + \ldots a_nx_n$$

and the n variables are independent,

$$\text{var } z = a_1^2 \text{ var } x_1 + a_2^2 \text{ var } x_2 + \ldots a_n^2 \text{ var } x_n. \qquad (9.63)$$

In particular we have, for the sampling variance of the difference of the means of two independent samples, say m_1' and p_1',

$$\text{var } (m_1' - p_1') = \frac{\mu_2}{n_1} + \frac{\varpi_2}{n_2}, \qquad (9.64)$$

μ_2 and ϖ_2 being the respective variances and n_1 n_2 the respective numbers in the samples.

Example 9.10

A random sample of 1,000 men from the North of England shows their mean wage to be 47 shillings a week with a standard deviation of 28 shillings. A random sample of 1,500 men from the South gives a mean wage of 49 shillings a week with a standard deviation of 40 shillings. Required to discuss the question whether the mean level of wages differs between North and South.

The difference of the means is 2 shillings and we wish to know whether this is significant.

From (9.64), taking as usual with large samples the unknown variances to be those of the samples, we find

$$\text{var (difference)} = \frac{28^2}{1000} + \frac{40^2}{1500}$$
$$= 1{\cdot}851.$$

The standard error is thus 1·36 and the difference in means, being less than twice this amount, is hardly significant of any real difference. Had the difference been three shillings instead of two we should probably have concluded that the difference, being more than twice the standard error, was significant.

There is an alternative approach to this problem which is worth noticing. Suppose we assume as our hypothesis under test that the distribution of wages in the two areas is the same. Then we may combine the sample figures to give a new estimate of the mean and variance in this distribution, e.g. the mean might be taken to be given by

$$\frac{(1000 \times 47) + (1500 \times 49)}{2500}$$
$$= 48{\cdot}2 \text{ shillings.}$$

In the first sample the sum of squares of deviations about the mean 47 is

$$1000 \times 28^2 = 784{,}000,$$

and hence the sum about the origin is $784{,}000 + (47^2 \times 1000) = 2{,}993{,}000$. Similarly in the second sample the sum of squares of deviations about the origin is

$$1500\ (40^2 + 49^2) = 6{,}001{,}500.$$

The second moment of the whole about the origin is then $\dfrac{8{,}994{,}500}{2500} = 3597{\cdot}8$, and hence the variance is $3597{\cdot}8 - (48{\cdot}2)^2 = 1274{\cdot}56$. We might take this as our estimate of the variance in the population and our problem would then be: does the mean in one of the parts of the whole sample, say the first, 47 shillings, differ significantly from the mean of the whole, 48·2 shillings?

Now at first sight it looks as if this is a case for the application of (9.64). We have two means, 47 and 48·2, with respective variances 784 and 1274·56, and require to know whether the means are significantly different. But the samples are no longer independent, for one of them is part of the other, and a modified formula must be used. If the means of the separate samples are $m_1'\left(= \dfrac{1}{n_1}\Sigma x_1\right)$ and $p_1'\left(= \dfrac{1}{n_2}\Sigma x_2\right)$ the mean of the two together is given by

$$\frac{n_1 m_1' + n_2 p_1'}{n_1 + n_2} = \frac{\Sigma x_1 + \Sigma x_2}{n_1 + n_2}$$

The difference of m_1' and this quantity, say q, is then

$$q = \frac{1}{n}\Sigma x_1 - \frac{\Sigma x_1 + \Sigma x_2}{n_1 + n_2}$$
$$= \frac{1}{n_1 + n_2}\left\{\frac{n_2}{n_1}\Sigma x_1 - \Sigma x_2\right\}.$$

Thus
$$E(q) = \frac{1}{n_1 + n_2}\left\{\frac{n_2}{n_1}.n_1\mu_1' - n_2\mu_1'\right\} = 0,$$

and hence

$$\text{var } q = E(q^2) = \frac{1}{(n_1 + n_2)^2} E\left\{\frac{n_2}{n_1}\Sigma x_1 - \Sigma x_2\right\}^2$$

Since x_1 and x_2 are independent, this reduces to

$$\frac{1}{(n_1 + n_2)^2} \quad n_1\mu_2 + n_2\mu_2 \Big]$$

$$\frac{n_2}{n_1(n_1 + n_2)}\mu_2.$$

In our case $n_1 = 1000$, $n_2 = 1500$ and our estimate of μ_2 is 1274·56. The variance of the difference then becomes, on substitution, 0·7647. The observed difference is $48{\cdot}2 - 47 = 1{\cdot}2$. Once again this is less than twice the standard error ($= \sqrt{\cdot 7647} = 0{\cdot}87$) and again we conclude that the difference is not significant.

REFERENCES

Davies, O. L., and Pearson, E. S. (1934), "Methods of estimating from samples the population standard deviation," *Supp. Jour. Roy. Stat. Soc.*, **1**, 76.

Fisher, R. A. (1920), "A mathematical examination of the methods of determining the accuracy of an observation by the mean error and the mean square error," *Monthly Notices Roy. Astr. Soc.*, **80**, 758.

—— and Tippett, L. H. C. (1928), "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. Camb. Phil. Soc.*, **24**, 180.

Gumbel, E. J. (1934), "Les valeurs extrêmes des distributions statistiques," *Annales de l'Institut Henri Poincaré*, **5**, 115.

Hartley, H. O. (1942), "The range in normal samples," *Biometrika*, **32**, 334.

Helmert (1876), *Astronomische Nachrichten*, **88**, No. 2096.

Hojo, T. (1931), "Distribution of the Median, Quartiles and Interquartile distance in samples from a normal population," *Biometrika*, **23**, 315.

Kondo, T. (1929), "On the standard error of the mean square contingency," *Biometrika*, **21**, 376.

Nair, U. S. (1936), "The standard error of Gini's mean difference," *Biometrika*, **28**, 428.

Pearson, E. S. (1926), "A further note on the distribution of range in samples taken from a normal population," *Biometrika*, **18**, 173.

—— and Adyanthaya, N. K. (1928), "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations," *Biometrika*, **20** A, 356.

—— (1932), "The percentage limits of the distribution of range in samples from a normal population," *Biometrika*, **24**, 404.

—— and Haines, Joan (1935), "The use of range in place of standard deviation in small samples," *Supp. Jour. Roy. Stat. Soc.*, **2**, 83.

—— and Hartley, H. O. (1942), "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, **32**, 301.

Pearson, K., and Filon, L. N. G. (1898), "On the probable errors of frequency constants and on the influence of random selection on variation and correlation," *Phil. Trans.*, **191**A, 229.

Pearson, K., "On the probable errors of frequency constants," Part I, *Biometrika*, 1903, **2**, 273; Part II, *Biometrika*, 1913, **9**, 1; Part III, *Biometrika*, 1920, 113.
—— (1913), "On the probable error of a coefficient of correlation as found from a four-fold table," *Biometrika*, **9**, 22.
—— (1915), "On the probable error of a coefficient of mean square contingency," *Biometrika*, **10**, 590.
—— (1931), "On the standard error of the median to a third approximation," *Biometrika*, **23**, 361.
—— and Pearson, M. V. (1931), "On the mean character and variance of a ranked individual and on the mean and variance of the intervals between ranked individuals," *Biometrika*, **23**, 364.
Tippett, L. H. C. (1925), "On the extreme individuals and the range of samples taken from normal population," *Biometrika*, **17**, 364.
Yasukawa, K. (1926), "On the probable error of the mode of skew frequency distributions," *Biometrika*, **18**, 263.

EXERCISES

9.1. Show that the mean value of the variance is given exactly by

$$E(m_2) = \frac{n-1}{n}\mu_2$$

and that its variance is given exactly by

$$\text{var}(m_2) = \left(\frac{n-1}{n}\right)^2 \frac{\mu_4 - \mu_2^2}{n} + \frac{2(n-1)}{n^2}\mu_2^2.$$

Hence verify that the formulae of this chapter as applied to the variance of a sample are accurate to order n^{-1}.

9.2. In the height distribution of Table 1.7 it has been found that

$$\begin{aligned} m_2 &= 6\cdot616 \\ m_3 &= -0\cdot207 \\ m_4 &= 137\cdot689. \end{aligned}$$

Regarding the distribution as a random sample from a population which is approximately normal, show that m_3 does not differ significantly from zero (which, of course, must be so if the assumption of normality is to be maintained) and that m_4 has a standard error of about 4 per cent. of its value.

9.3. Verify that the standard error of the first decile in samples from a normal population is $\dfrac{1\cdot709\sigma}{n}$.

9.4. In the distribution of Australian marriages of Table 1.8 it has been found that the mean is 29·4 years, the standard deviation 8 years approximately. The median frequency is about 63,150. Taking this distribution to be a random sample, show that the standard error of the mean is 0·015 years and that of the median 0·043 years.

9.5. If a series of random samples of different sizes is drawn from a population in which the proportion of members bearing an attribute A is ϖ, show that the variance of the proportions of A in such sets is $\dfrac{\varpi(1-\varpi)}{H}$ where H is the harmonic mean of the numbers in the samples.

9.6. Show that the sampling variances of the first four cumulants, as calculated from the moments, are given to order n^{-1} by

$$\text{var}\,\kappa_1 = \frac{1}{n}\kappa_2$$

$$\text{var}\,\kappa_2 = \frac{1}{n}(\kappa_4 + 2\kappa^2)$$

$$\text{var}\,\kappa_3 = \frac{1}{n}(\kappa_6 + 9\kappa_4\kappa_2 + 9\kappa_3^2 + 6\kappa_2^3)$$

$$\text{var}\,\kappa_4 = \frac{1}{n}(\kappa_8 + 16\kappa_6\kappa_2 + 48\kappa_5\kappa_3 + 34\kappa_4^2 + 72\kappa_4\kappa_2^2 + 144\kappa_3^2\kappa_2 + 24\kappa_2^4).$$

9.7. If the variate range is divided into sub-ranges and the frequency of a large sample falling into the pth range is f_p, show that

$$\text{var}\,f_p = f_p\left(1 - \frac{f_p}{n}\right)$$

$$\text{cov}\,(f_p, f_q) = -\frac{1}{n}f_p f_q$$

and hence find expressions for the sampling variance of the rth moment about an arbitrary point.

9.8. Show that in odd samples of n from a rectangular population of unit range the sampling variance of the distribution of the median is given exactly by $\dfrac{1}{4(n+2)}$.

CHAPTER 10

EXACT SAMPLING DISTRIBUTIONS

10.1. The role of the sampling distribution in statistical inference has been indicated in Chapter 8. In the present chapter we propose to give an account of the main methods of finding such distributions when the population from which the sample was derived is specified. It will, as usual, be assumed that the sampling is simple and random. Thus, if the parent distribution is $dF(x)$ the simultaneous distribution of n values $x_1 \ldots x_n$ is $dF(x_1)\, dF(x_2) \ldots dF(x_n)$; and if z is a statistic

$$z = z(x_1 \ldots x_n) \quad . \qquad (10.1)$$

the distribution function of z is given by

$$F(z) = \int \ldots \int dF(x_1) \ldots dF(x_n) \quad . \qquad (10.2)$$

the integration being taken over the domain of the x's such that $z(x_1 \ldots x_n) \leqslant z_0$.

Formally, (10.2) is the solution of our problem, which thus reduces to the purely mathematical one of evaluating certain multiple integrals or sums. The methods with which we are here concerned are fundamentally devices of various kinds to facilitate the integrative process. They may be classified into four groups:—

(*a*) straightforward evaluation of the integral (10.2) by ordinary analytical processes such as a convenient change of variable;

(*b*) the use of geometrical terminology to effect the same object and to avoid cumbrous analytical formulae ;

(*c*) the use of characteristic functions; and

(*d*) other analytical methods, including mathematical induction.

10.2. As an illustration of the straightforward analytical approach, let us find the distribution of the sums of squares of n variables, each of which is distributed normally with unit variance and zero mean. The joint distribution of the n variables is then the product of n quantities of type $\frac{1}{\sqrt{(2\pi)}} e^{-\frac{x^2}{2}}$, that is to say

$$dF = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp - \tfrac{1}{2}(x_1^2 + x_2^2 + \ldots x_n{}^2)\, dx_1 \ldots dx_n . \qquad (10.3)$$

We require the sampling distribution of

$$z = x_1^2 + x_2^2 + \ldots x_n{}^2 . \qquad (10.4)$$

We have thus to evaluate the multiple integral

$$dF = \int \ldots \int \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\tfrac{1}{2}\Sigma x^2\right) dx_1 \ldots dx_n$$

over the domain of x's conditioned by (10.4).

Make the transformation to variables z, θ_1, θ_2, . . . θ_{n-1}

$$\begin{aligned} x_1 &= z^{\frac{1}{2}} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} \\ x_2 &= z^{\frac{1}{2}} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-2} \sin\theta_{n-1} \\ x_j &= z^{\frac{1}{2}} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-j} \sin\theta_{n-j+1} \\ x_n &= z^{\frac{1}{2}} \sin\theta_1 \end{aligned} \quad . \quad (10.5)$$

The Jacobian of this transformation is given by

$$\frac{\partial(x_1, \ldots x_n)}{\partial(z, \theta_1, \ldots \theta_{n-1})},$$

which is equal to $\frac{1}{2}z^{\frac{n-2}{2}}$ times the determinant

$$\begin{array}{cccc} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} & \cos\theta_1 \cos\theta_2 & .\ \cos\theta_{n-2} \sin\theta_{n-1} & \sin\theta_1 \\ -\sin\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} & -\sin\theta_1 \cos\theta_2 & .\ \cos\theta_{n-2} \sin\theta_{n-1} & \cos\theta_1 \\ -\cos\theta_1 \sin\theta_2 \ldots \cos\theta_{n-1} & -\cos\theta_1 \sin\theta_2 & .\ \cos\theta_{n-2} \sin\theta_{n-1} & 0 \\ & & & \\ -\cos\theta_1 \cos\theta_2 \ldots \sin\theta_{n-1} & +\cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} & & 0 \end{array}$$

Taking out common factors in columns we find that this determinant is equal to
$\cos^{n-1}\theta_1 \cos^{n-2}\theta_2 \ldots \cos\theta_{n-1} \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-1}$ times

$$\begin{array}{ccccc} & & & & 1 \\ -\tan\theta_1 & -\tan\theta_1 & -\tan\theta_1 & & \cot\theta_1 \\ -\tan\theta_2 & -\tan\theta_2 & -\tan\theta_2 & & 0 \\ \ldots & \ldots & \ldots & & \\ -\tan\theta_{n-2} & -\tan\theta_{n-2} & \cot\theta_{n-2} & 0 & \\ -\tan\theta_{n-1} & \cot\theta_{n-1} & 0 & 0 & \end{array}$$

and, on subtracting each column from the preceding one, the determinant is found to reduce to $\cos^{n-2}\theta_1 \cos^{n-3}\theta_2 \ldots \cos\theta_{n-2}$.

Thus our integral becomes

$$\int \ldots \int \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}z} \tfrac{1}{2} z^{\frac{n-2}{2}} \cos^{n-2}\theta_1 \ldots \cos\theta_{n-2}\, dz\, d\theta_1 \ldots d\theta_{n-1} \quad . \quad (10.6)$$

The advantage of the transformation is that the limits of the variables are now much simpler. z itself can vary from 0 to z and the θ's from 0 to 2π. Thus the integral (10.6) divides into a product of integrals, those in θ being constant, and we find for our distribution function of z

$$F(z) = k\int_0^z e^{-\frac{1}{2}z} z^{\frac{n-2}{2}}\, dz. \quad . \quad (10.7)$$

The constant k may be evaluated by integration between 0 and ∞ and we have

$$\begin{aligned} 1 &= k\int_0^\infty e^{-\frac{1}{2}z} z^{\frac{n-2}{2}}\, dz \\ &= k2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right). \end{aligned}$$

Hence the distribution sought is

$$dF = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}z} z^{\frac{n-2}{2}} dz, \qquad \infty \qquad . \quad (10.8)$$

a Pearson Type III curve.

10.3. The essential feature of the change of variables is the simplification of the domain of integration as defined by the limits of the new variables. In general, we usually take the statistic whose sampling distribution is being sought to form one of the new variables and choose $n - 1$ others in any way which may be convenient to the particular problem. Then, if J is the Jacobian of the transformation, namely

$$J = \frac{\partial(x_1 \ldots x_n)}{\partial(z, \theta_1, \ldots \theta_{n-1})},$$

the integral (10.2) becomes

$$F(z) = \int \ldots \int f(x_1) \ldots f(x_n) \frac{\partial(x_1, \ldots x_n)}{\partial(z, \theta_1, \ldots \theta_{n-1})} dz\, d\theta_1 \ldots d\theta_{n-1}, \quad . \quad (10.9)$$

$f(x_j)$ being the frequency function of the parent and x_j being expressed in terms of z and the θ's. The integration now takes place with respect to the θ's, which can usually be chosen so as to vary between limits which are independent of z; and thus the indefinite integral (10.2) is replaced by more easily calculable definite integrals.

As always in such cases J is subject to an ambiguity of sign which must be determined so as to make the transformed integral positive. The validity of the variate-transformation depends on the familiar conditions governing the change of variable in a multiple integral. For example, it is a sufficient condition that the new variables and their first derivatives shall be continuous in the x's and that J does not change sign in the domain of integration.* Some further examples will make the general type of investigation clear.

Example 10.1

To find the distribution of the mean of a sample of n values $x_1 \ldots x_n$ from the distribution

$$dF = \frac{dx}{\pi(1 + x^2} \qquad -\infty \leqslant x \leqslant \infty.$$

The joint distribution is

$$\frac{1}{\pi^n} \prod_{j=1}^{n} \frac{dx_j}{(1 + x_j^2)} \qquad (10.10)$$

and the statistic z is given by

$$nz = \sum_{j=1} x_j. \qquad . \quad (10.11)$$

We have to integrate (10.10) over a domain of x's subject to $\Sigma x \leqslant nz$. Let us take new variables $x_1 = x_1$, $x_2 = x_2$, $\ldots$ $x_{n-1} = x_{n-1}$ and

$$x_n = nz - x_1 - x_2 - \ldots x_{n-1}.$$

* See, for example, de la Vallée Poussin, *Cours d'analyse infinitésimal*, 1926, vol. 1, para. 285; vol. 2, para. 18.

Here J is evidently equal to the constant n. Our new variables $x_1 \ldots x_{n-1}$ may extend from $-\infty$ to $+\infty$ and the new variable z from $-\infty$ to z. We then have

$$F(z) = \int_{-\infty}^{z} dz \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{n}{\pi^n} \prod_{j=1}^{n-1} \frac{dx_j}{(1+x_j^2)\{1+(nz-x_1-\ldots-x_{n-1})^2\}} \quad . \quad (10.12)$$

and the *frequency* function of z is given by the $(n-1)$-fold multiple integral in $x_1 \ldots x_{n-1}$ in (10.12). This integral may be evaluated by step-by-step integration. We have

$$\frac{1}{(1+x^2)\{r^2+(a-x)^2\}} = \frac{1}{\{a^2+(r+1)^2\}\{a^2+(r-1)^2\}} \left[\frac{2ax}{x^2+1} + \frac{a^2+r^2}{x^2+1} + \frac{2a^2-2ax}{r^2+(a-x)^2} + \frac{a^2-r^2+1}{r^2+(a-x)^2}\right].$$

Whence, integrating with respect to x from $-\infty$ to $+\infty$, we find on the right

$$\frac{1}{\{a^2+(r+1)^2\}\{a^2+(r-1)^2\}}\Big[a\log(x^2+1) - a\log\{r^2+(a-x)^2\}$$
$$+ (a^2+r^2-1)\tan^{-1}x + \frac{r^2+1}{r}\tan^{-1}\frac{x-a}{r}\Big]_{-\infty}^{\infty}$$

reducing to
$$\pi\left(\frac{r+1}{r}\right)\left\{\frac{1}{a^2+(r+1)^2}\right\} \quad . \quad (10.13)$$

Thus in (10.12), taking $x = x_{n-1}$, $r = 1$, $a = nz - x_1 - \ldots - x_{n-2}$, we find that the $(n-1)$-fold integral reduces to

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{n}{\pi^{n-1}} \prod_{j=1} \frac{dx_j}{(1+x_j^2)\{2^2+(nz-x_1-\ldots-x_{n-2})^2\}}.$$

Integrating with respect to $x_{n-2}, x_{n-1} \ldots$ successively, we reduce this eventually to

$$\frac{n^2}{\pi\{n^2+(nz)^2\}} = \frac{1}{\pi(1+z^2)} \quad . \quad (10.14)$$

Thus the distribution of z is given by

$$dF = \frac{dz}{\pi(1+z^2)} \qquad -\infty \leqslant z \leqslant \infty \quad . \quad (10.15)$$

and is thus the same as that of a single observation.

This is an interesting example of the failure of the Central Limit Theorem, the mean of samples of n failing to tend to normality for large n. The second moment of the distribution does not exist.

Example 10.2

To find the distribution of a linear function of n variables $x_1 \ldots x_n$ where x_j is distributed normally with zero mean and variance v_j.

Let the linear function be

$$z = a_1x_1 + \ldots + a_nx_n. \quad . \quad (10.16)$$

Then by a transformation $\xi_j = \frac{x_j}{\sqrt{v_j}}$ we have

$$z = \Sigma a_j\sqrt{v_j}\xi_j \quad . \quad (10.17)$$

and ξ_j is now distributed with zero mean and unit variance. Our problem is thus equivalent to finding the distribution of a linear function of variables each of which is normally distributed with zero mean and unit variance.

Consider a transformation of type

$$\left.\begin{aligned}\zeta_1 &= l_1\xi_1 + l_2\xi_2 + \ldots + l_n\xi_n\\ \zeta_2 &= m_1\xi_1 + m_2\xi_2 + \ldots + m_n\xi_n\\ &\ldots\ldots\ldots\\ \zeta_n &= p_1\xi_1 + p_2\xi_2 + \ldots + p_n\xi_n\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (10.18)$$

and let us determine the l's . . . p's such that

$$\begin{aligned}&l_j^2 + m_j^2 + \ldots p_j^2 = 1, && \text{all } j\\ &l_jl_k + m_jm_k + \ldots p_jp_k = 0, && \text{all } j, k, \quad j \neq k\end{aligned} \quad . \quad (10.19)$$

This can always be done, for the conditions impose only $n + \frac{1}{2}n(n-1)$ conditions on the n^2 constants.

We have then

$$\sum_{j=1}^{n}\zeta_j^2 = (l_1\xi_1 + \ldots + l_n\xi_n)^2 + \ldots + (p_1\xi_1 + \ldots + p_n\xi_n)^2 = \sum_{j=1}^{n}\xi_j^2$$

in virtue of (10.19). The joint distribution of the ξ's is by hypothesis

$$\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\tfrac{1}{2}\Sigma\xi^2\right)\Pi\, d\xi$$

$$\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\tfrac{1}{2}\Sigma\zeta^2\right)J\,\Pi\, d\zeta \qquad (10.20)$$

where

$$J = \left|\frac{\partial\xi}{\partial\zeta}\right| = \frac{1}{\left|\frac{\partial\zeta}{\partial\xi}\right|}.$$

The determinant $\frac{1}{J}$ is then, from (10.18),

$$\begin{vmatrix} l_1 & l_2 & \ldots & l_n\\ m_1 & m_2 & \ldots & m_n\\ \ldots & & & \\ p_1 & p_2 & \ldots & p_n\end{vmatrix}$$

and multiplying this by the equal determinant

$$\begin{vmatrix} l_1 & m_1 & \ldots & p_1\\ l_2 & m_2 & \ldots & p_2\\ & & \ldots & \\ l_n & m_n & \ldots & p_n\end{vmatrix}$$

we find, in virtue of (10.19), that the product is

$$\begin{vmatrix} 1 & 0 & \ldots & 0\\ 0 & 1 & \ldots & 0\\ & & \ldots & \\ 0 & 0 & \ldots & 1\end{vmatrix} = 1.$$

Thus $\frac{1}{J} = \pm 1$ and (10.20) becomes

$$\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\tfrac{1}{2}\Sigma\zeta^2\right)\Pi\, d\zeta. \qquad . \quad (10.21)$$

Now the ξ's may vary from $-\infty$ to ∞, and if we require the distribution of one of the ζ's, say ζ_1 ($= l_1\xi_1 + \ldots l_n\xi_n$), we have to integrate over all values of ξ such that $\Sigma l_j\xi_j \leqslant \zeta_1$. This is equivalent to a range of ζ_1 from $-\infty$ to ζ_1 and of the other ζ's from $-\infty$ to $+\infty$. Thus the integral of (10.21) becomes the product of $(n-1)$ definite integrals each equal to $\int_{-\infty}^{\infty} e^{-\frac{1}{2}\zeta^2}\,d\zeta = (2\pi)^{\frac{1}{2}}$ and the integral $\int_{-\infty}^{\zeta} e^{-\frac{1}{2}\zeta^2}\,d\zeta$, and hence reduces to

$$F(\zeta) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\zeta} e^{-\frac{1}{2}\zeta^2}\,d\zeta. \tag{10.22}$$

In other words, ζ is distributed normally with unit variance and zero mean. ζ is an arbitrary linear function $\Sigma l_j\xi_j$ subject to the condition that $\Sigma l_j^2 = 1$. Referring to (10.17) we see that the slightly more general linear function $z = \Sigma a_j x_j = \Sigma a_j\sqrt{v_j}\xi_j$ will be distributed normally about zero mean with variance $\Sigma a_j^2 v_j$, for then $\dfrac{\Sigma a_j\sqrt{v_j}\xi_j}{\Sigma a_j^2 v_j}$ has coefficients $l_j\left(=\dfrac{a_j\sqrt{v_j}}{\Sigma a_j^2 v_j}\right)$ obeying the condition $\Sigma l_j^2 = 1$ and is distributed with unit variance.

The Geometrical Method

10.4. A considerable amount of cumbrous analysis may usually be avoided by the use of geometrical representation of the domain of integration. We may imagine the values $x_1 \ldots x_n$ attaching to any given sample as the co-ordinates of a point in an n-dimensional Euclidean hyperspace. The function $dF(x_1) \ldots dF(x_n)$ may then be regarded as the *density* at the point and the total frequency between z_1 and z_2 will be the integral of this density (the *weight*) in a region lying between the two loci $z(x_1 \ldots x_n) = z_1$ and $z(x_1 \ldots x_n) = z_2$, which in general will be hypersurfaces in the n-fold space, i.e. will themselves be spaces of $(n-1)$ dimensions. The distribution function of z will be the total weight between the hypersurface corresponding to $z = -\infty$ and that corresponding to z; and the frequency function will be the element of weight between the hypersurfaces $z - \frac{1}{2}dz$ and $z + \frac{1}{2}dz$.

Example 10.3

Consider again the problem of Example 10.2. In the n-fold ξ-space the density is given by

$$\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\tfrac{1}{2}\Sigma\xi^2\right).$$

The statistic z $(= \Sigma a_j x_j)$ determines a hyperplane

$$z = \Sigma a_j\sqrt{v_j}\xi_j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{10.23}$$

and we have to find the total weight between this hyperplane and the corresponding hyperplane at $-\infty$, i.e. the weight on one side—the " lower " side—of the hyperplane (10.23).

Now $\Sigma\xi^2$ is the square of the distance of the point $\xi_1 \ldots \xi_n$ from the origin and is therefore unchanged by any rotation of the co-ordinate axes. Choose such a rotation which brings the axis of one variable perpendicular to the hyperplane (10.23), meeting it in Q. Let P be the sample point $\xi_1 \ldots \xi_n$ and O the origin. Then

$$\Sigma\xi^2 = OP^2 = OQ^2 + QP^2,$$

so that the density at P is

$$\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}OQ^2} e^{-\frac{1}{2}QP^2}.$$

For variation over the hyperplane OQ^2 is constant and the integral of $e^{-\frac{1}{2}QP^2}$ is thus a constant independent of OQ. Hence the frequency function of z is given by

$$f(z) = ke^{-\frac{1}{2}OQ^2},$$

k being some constant.

But OQ is the distance from O to the hyperplane and is given by

$$OQ^2 = \frac{z^2}{\Sigma a_j^2 v_j}.$$

Hence

$$f(z) = k \exp -\frac{1}{2}\frac{z^2}{\Sigma a_j^2 v_j},$$

i.e. z is distributed normally with variance $\Sigma a_j^2 v_j$ about zero mean.

The reader will find it instructive to compare this example with the previous one. They are, in effect, the same thing expressed in different language.

Example 10.4

Consider again the illustration of **10.2**. The elegance of the geometrical approach is well brought out by the analogous derivation of the result there obtained.

In fact, our density function, as before, is given by

$$ke^{-\frac{1}{2}OP^2}.$$

We require the distribution of the statistic $z = OP^2$, and the density is obviously constant over the surface z = constant, that is to say the $(n-1)$-dimensional hypersphere. The frequency function of z is then the integral of this constant density between the hyperspheres z and $z + dz$, i.e. is proportional to $e^{-\frac{1}{2}OP^2}$ times the element of the volume of the hypersphere, which itself is proportional to the nth power of the radius OP. Thus we have

$$\begin{aligned} dF &= ke^{-\frac{1}{2}OP^2}\frac{d}{dz}OP^n\, dz \\ &= ke^{-\frac{1}{2}z}\, z^{\frac{1}{2}(n-2)}\, dz, \end{aligned}$$

giving, on evaluation of the constant,

$$dF = \frac{1}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-2)}\, dz$$

as before.

Now suppose that the quantities $x_1 \ldots x_n$, while still being normally distributed with unit variance, are subject to p linear restrictions of type

$$a_1x_1 + a_2x_2 + \ldots a_nx_n = b.$$

In the n-space the variables x will then be constrained to lie on p hyperplanes. The first will cut the hypersphere of constant density in a hypersphere of one lower dimension, also, of course, of constant density; the second will cut this in a hypersphere of one lower dimension still, and so on. The result of the linear restrictions will be to constrain the

variables to a hypersphere of p lower dimensions, and thus the distribution of z in these circumstances will be as before, but with $n - p$ instead of n, i.e.

$$dF = \frac{e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-p-2)} dz}{2^{\frac{1}{2}(n-p)} \Gamma\left(\frac{n-p}{2}\right)} \qquad (10.24)$$

Example 10.5. The sampling distribution of the mean and variance in normal samples

Writing $\bar{x}$ for the mean of a sample, we have, for the variance s^2,

$$s^2 = \frac{1}{n}\Sigma(x - \bar{x})^2$$

$$= \frac{1}{n}\Sigma x^2 - \bar{x}^2.$$

In samples from a normal population with zero mean and unit variance the density at the point $x_1 \ldots x_n$ is proportional to

$$\exp\ (-\tfrac{1}{2}\Sigma x^2) = \exp\ \{-\tfrac{1}{2}(ns^2 + n\bar{x}^2)\}. \quad . \quad . \quad . \quad (10.25)$$

Let us find the sampling distributions of s and $\bar{x}$. From (10.25) it is seen that the density function can be expressed simply in terms of those quantities, and we then have to find some transformation of the volume element $dx_1 \ldots dx_n$.

In the n-space consider the unit vector whose direction cosines are $\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \quad \frac{1}{\sqrt{n}}$, say OQ where O is the origin. If P is the sample point, let PM be the perpendicular from P on to OQ. Then the length of OM is

$$\frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \quad \frac{x_n}{\sqrt{n}} = \bar{x}\sqrt{n}.$$

The length of OP is $\sqrt{\Sigma x^2}$. Thus the length of PM is $(\Sigma x^2 - n\bar{x}^2)^{\frac{1}{2}} = s\sqrt{n}$.

The element of volume at P may be regarded as the product of an elemental increment in OM, equal to $d\bar{x}$, and the elemental volume in the perpendicular hyperplane through M. In the hyperplane the contours of equal density, as in the last example, are hyperspheres of radius $s\sqrt{n}$ centred at M, and consequently the element of volume is equal to $k\, d\bar{x}\, s^{n-2}\, ds$ multiplied by other elements which need not concern us since they are independent of $\bar{x}$ and s. We have then for the element of frequency

$$dF \propto \exp\ \{-\tfrac{1}{2}(ns^2 + n\bar{x}^2)\}s^{n-2}\, d\bar{x}\, ds \quad . \quad . \quad . \quad . \quad (10.26)$$

and this splits into two factors

$$dF \propto e^{-\frac{1}{2}n\bar{x}^2} d\bar{x} \qquad (10.27)$$

$$dF \propto e^{-\frac{1}{2}ns^2} s^{n-2} \qquad (10.28)$$

Thus in samples from a normal population the distributions of mean and variance are independent. Equation (10.27) is equivalent to the result found in Examples 10.2 and 10.3. Equation (10.28) is new. We have

$$dF \propto e^{-\frac{1}{2}ns^2} s^{n-3}\, ds^2$$

and, on evaluation of the constant,

$$dF = \frac{n^{\frac{n-1}{2}}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{1}{2}ns^2} s^{n-3}\, d(s^2), \qquad 0 \leqslant s \leqslant \infty. \quad . \quad (10.29)$$

It is interesting to compare this with the distribution of the previous example. In the latter case we found the distribution of the sum of squares of the variables *measured from a fixed point*. In this case we have found the distribution of $\frac{1}{n}$th of the sum of the squares *measured from the sample mean*. A comparison of the form (10.29) with that of (10.24) shows that the distribution of variances is, except for constants, the same as that of sums of squares when subject to one linear constraint.

Example 10.6. "Student's" distribution

In the previous example we have

$$\frac{\bar{x}\sqrt{n}}{s\sqrt{n}} = \frac{OM}{PM} = \cot\phi,$$

where ϕ is the angle POM.

If, then, we define a statistic $z = \frac{\bar{x}}{s}$, z will be constant over the cone obtained by rotating PO about the unit vector, keeping the angle ϕ constant. The distribution of z will then be given by determining the weight between the cones defined by ϕ and $\phi + d\phi$.

Consider the intersection of these cones with the hypersphere of radius OP. They will cut off an annulus on the sphere whose "content" (the n-dimensional analogue of volume) will be proportional to $OP\,d\phi . PM^{n-2}$

$$= OP^{n-1}\sin^{n-2}\phi\, d\phi.$$

The density function is constant and proportional to $e^{-\frac{1}{2}OP^2}$ on the hypersphere and thus the total frequency between the cones will be proportional to

$$\int_0^\infty e^{-\frac{1}{2}OP^2} OP^{n-1}\sin^{n-2}\phi\, d\phi\, d(OP)$$

$$\propto \sin^{n-2}\phi\, d\phi, \qquad 0 \leqslant \phi \leqslant \pi.$$

The distribution of z $(= \cot\phi)$ is then given by

$$dF \propto \frac{k\,dz}{(1+z^2)^{\frac{n}{2}}}$$

or, on evaluation of the constant,

$$dF = \frac{dz}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)(1+z^2)^{\frac{n}{2}}} \tag{10.30}$$

Since z is the ratio of two functions of the variables of unit dimension this distribution holds for samples from a normal population irrespective of the scale, that is to say, irrespective of the variance of the parent population.

The distribution is usually put in a slightly different form.

Put
$$t = \frac{\bar{x}\sqrt{n}}{\left\{\frac{1}{n-1}\Sigma(x-\bar{x})^2\right\}^{\frac{1}{2}}} = \sqrt{(n-1)}\,z.$$

(10.30) then becomes

$$dF = \frac{1}{\sqrt{(n-1)}B\left(\frac{n-1}{2}, \frac{1}{2}\right)} \frac{dt}{\left(1+\frac{t^2}{n-1}\right)^{\frac{n}{2}}} \quad . \quad . \quad . \quad (10.31)$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \frac{dt}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \quad . \quad (10.32)$$

where $\nu = n - 1$.

This celebrated expression is known as "Student's" distribution after the *nom de plume* of its discoverer (1908).* The distribution function may be evaluated from the incomplete B-function, but special tables have been prepared. One such, due to "Student" himself, is given as Appendix Table 3.

Example 10.7. Distribution of the mean of samples from a rectangular population

Consider now a sample of n values from the rectangular distribution

$$dF = dx \qquad 0 \leqslant x \leqslant 1.$$

In the n-space the density function will be a constant everywhere inside a hypercube

$$0 \leqslant x_j \leqslant 1, \quad j = 1, \quad . \; . \; . \; n \quad . \quad . \quad . \quad . \quad (10.33)$$

and zero elsewhere. The unit vector will be the long diagonal of this cube. If P is the sample point $(x_1 \; . \; . \; . \; x_n)$ and PM the perpendicular on to this diagonal, then, as shown in Example 10.5, $OM = \bar{x}\sqrt{n}$. Thus, for the distribution of $\bar{x}$ we require the element of weight (which in this case is proportional to the element of volume) between the hyperplanes $\bar{x}$ and $\bar{x} + d\bar{x}$; and this is equivalent to finding the content of the hyperplane (its "area") cut off by the various faces of the hypercube. The complication of the problem arises from the fact that as $\bar{x}$ increases this region changes its shape according to the number of edges of the hypercube cut by the hyperplane.

Consider the "quadrants"

$$\begin{matrix} x_j \geqslant r_j \\ r_j = 0 \text{ or } 1 \end{matrix} \quad j = 1, \; 2, \qquad n, \qquad (10.34)$$

whose corners are the corners of the hypercube. Any one of the corners may have 0 or 1 or 2 . . . or n of its co-ordinates equal to unity and the rest zero. We divide the quadrants into $(n + 1)$ sets according as the corner has 0, 1, . . . n of its co-ordinates equal to unity, that is, according as

$$= \sum_{j=1}^{n} r_j$$

is equal to 0, 1, . . . n. A quadrant of the tth set may be called Q_t. There will be $\binom{n}{t}$ different Q_t's. Let S be any point of Q_0, i.e. any point whose co-ordinates are all $\geqslant 0$,

* Strictly speaking, "Student's" distribution is that of (10.30), the modified form (10.32) being due to R. A. Fisher. The latter form is therefore sometimes referred to as Fisher's t-distribution.

and let just s of its co-ordinates be $\geqslant 1$. Then S will belong to just $\binom{s}{0} = 1$, Q_0; $\binom{s}{1}Q_1$'s; Q_2's, and so on. Now if $s > 0$,

$$\sum_{t=0}^{s}(-1)^t\binom{s}{t} = (1-1)^s \quad 0. \qquad . \quad (10.35)$$

Hence, if whenever a point belongs to a Q_t we give it a density $(-1)^t$ and then sum over all Q, the resultant density will be 1 or 0 according as the point belongs to the hypercube or not.

Let the segment of the hyperplane

$$z = \Sigma(x) \quad . \quad . \quad . \quad . \quad . \quad . \quad (10.36)$$

lying in Q_0 have content $V_n(z)$. Then the segment lying in any member of (10.34) will have content $V_n(z-r)$ which is zero if $r \geqslant z$. Further, the segment of (10.36) lying in any member of (10.34) will have the content

$$\sum_{r=0}^{k}(-1)^r\binom{n}{r}V_n(z-r) \quad . \quad . \quad . \quad (10.37)$$

where $k = [z]$, = the greatest integer less than z.

To find $V_n(z)$, let $\bar{V}_{n-1}(z)$ be the projection of $V_n(z)$ perpendicular to one of the axes, so that

$$V_n(z) = \sqrt{n}\,\bar{V}_{n-1}(z).$$

Now $\bar{V}_n(z)$ is the content of the n-dimensional region bounded by (10.36) and the co-ordinate hyperplanes—a region whose base is therefore of content $V_n(z)$. The perpendicular from O to this base is $\frac{z}{\sqrt{n}}$. Hence

$$\bar{V}_n(z) = \frac{1}{n}\left(\frac{z}{\sqrt{n}}\right)V_n(z)$$

and

$$\bar{V}_{n-1}(z) = \frac{1}{n-1}\left(\frac{z}{\sqrt{(n-1)}}\right)V_{n-1}(z)$$

or

$$V_n(z) = \frac{1}{n-1}\sqrt{\frac{n}{n-1}}\,zV_{n-1}(z). \quad . \quad . \quad . \quad (10.38)$$

Since $V_2(z) = z\sqrt{2}$ repeated applications of this formula give

$$V_n(z) = \frac{\sqrt{n}}{(n-1)!}z^{n-1}.$$

Substituting in (10.37) we find for the content of the region common to the hypercube and the hyperplane

$$f(z) = \frac{\sqrt{n}}{(n-1)!}\sum_{r=0}^{k}(-1)^r\binom{n}{r}(z-r)^{n-1} \quad . \quad . \quad . \quad . \quad (10.39)$$

for values of z between k and $k+1$.

Since

$$\int_0^n f(z)\frac{dz}{\sqrt{n}} =$$

the distribution of the mean $m = \frac{z}{n}$ is given by

$$f(m) = \frac{n^n}{(n-1)!}\sum_{r=0}^{k}(-1)^r\binom{n}{r}\left(m - \frac{r}{n}\right)^{n-1} dm \qquad \frac{k}{n} \leqslant m \leqslant \frac{k+1}{n}. \tag{10.40}$$

This is the required distribution. It is unusual in consisting of n arcs of degree $(n - 1)$ in m, having $(n - 1)$-point contact at their joins, that is at the points $\frac{k}{n}$ $(k = 1, 2, \ldots n)$.

The distribution is symmetrical since the hyperplane z = constant is perpendicular to the long diagonal, which itself is an axis of symmetry of the hypercube.

For particular values $n = 2, 3, 4$, (10.40) gives the following results for the frequency function :—

$n = 2$:	$4m$,	$0 \leqslant m \leqslant \frac{1}{2}$
	$4(1 - m)$,	$\frac{1}{2} \leqslant m \leqslant 1$
$n = 3$:	$\frac{27m^2}{2}$,	$0 \leqslant m \leqslant \frac{1}{3}$
	$\frac{27}{2}\{m^2 - 3(m - \frac{1}{3})^2\}$,	$\frac{1}{3} \leqslant m \leqslant \frac{2}{3}$
	$\frac{27}{2}(1 - m)^2$,	$\frac{2}{3} \leqslant m \leqslant 1$
$n = 4$	$\frac{128}{3}m^3$,	$0 \leqslant m \leqslant \frac{1}{4}$
	$\frac{128}{3}\{m^3 - 4(m - \frac{1}{4})^3\}$,	$\frac{1}{4} \leqslant m \leqslant \frac{1}{2}$
	$\frac{128}{3}\{(1 - m)^3 - 4(\frac{3}{4} - m)^3\}$,	$\frac{1}{2} \leqslant m \leqslant \frac{3}{4}$
	$\frac{128}{3}(1 - m)^3$,	$\frac{3}{4} \leqslant m \leqslant 1.$

If the frequency curve be drawn it will be found to resemble a normal curve in appearance. The distribution, of course, tends to normality as n increases in virtue of the Central Limit Theorem.

The Method of Characteristic Functions

10.5. It has already been noted that the characteristic function of the sum of n independent variables is the product of their characteristic functions. This simple property enables us to find the sampling distribution of a wide class of statistics which are expressible as sums, and particularly of the mean.

If we have a sample of n values from a population whose characteristic function is $\phi(t)$, the characteristic function of their sum is ϕ^n. Thus the distribution function of their sum z is

$$F(z) - F(0) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{1 - e^{-ixt}}{it}\phi^n \, dt \tag{10.41}$$

and the frequency function is

$$f(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\phi^n \, dt. \tag{10.42}$$

The following examples will illustrate the power of these results.

√ *Example 10.8. Distribution of the Mean for the Binomial*

The characteristic function of the binomial $(q + p)^r$ is

$$(q + pe^{it})^r.$$

The c.f. of the sampling distribution of the sum of n values is then

$$(q + pe^{it})^{rn}$$

and that of the distribution of the mean $\left(\frac{1}{n} \text{ of that sum}\right)$ is

$$\left(q + pe^{\frac{it}{n}}\right)^{rn}.$$

But this is the c.f. of the binomial

$$(q + p)^{rn}, \tag{10.43}$$

the interval being $\frac{1}{n}$ instead of unity; and hence this distribution is that of the mean.

√ *Example 10.9. Distribution of the Mean for the Poisson Distribution*

The characteristic function of the Poisson distribution whose general term is $e^{-\lambda}\frac{\lambda^r}{r!}$ is

$$\exp\{\lambda(e^{it} - 1)\}.$$

The c.f. of the mean is then

$$\exp n\lambda\left(e^{\frac{it}{n}} - 1\right)$$

and hence the distribution of the mean is the Poisson distribution, whose general term is

$$e^{-n\lambda}\frac{(n\lambda)^r}{r!}, \quad . \quad . \quad . \quad . \quad . \tag{10.44}$$

the interval being $\frac{1}{n}$ instead of unity.

√ *Example 10.10. Distribution of the Mean for the Normal Population*

The characteristic function of the normal distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}dx$$

is

$$\exp\{-\tfrac{1}{2}t^2\sigma^2 + it\mu\}.$$

The c.f. of the distribution of the mean of n values is then

$$\exp n\left\{-\tfrac{1}{2}\frac{t^2\sigma^2}{n^2} + \frac{it\mu}{n}\right\} = \exp\left\{-\tfrac{1}{2}\frac{t^2\sigma^2}{n} + it\mu\right\} \tag{10.45}$$

This is the c.f. of a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, which is therefore the distribution required.

Example 10.11. Distribution of the Mean for the Type III Population

The characteristic function of the distribution

$$dF = \frac{1}{\Gamma(\gamma)} e^{-\frac{x}{a}} \left(\frac{x}{a}\right)^{\gamma-1} \frac{dx}{a}, \qquad a > 0$$

is

$$\frac{1}{(1 - ita)^{\gamma}}.$$

The c.f. of the distribution of the mean of n values is then

$$\frac{1}{\left(1 - \frac{ita}{n}\right)^{n\gamma}}.$$

This is the c.f. of the distribution

$$dF = \frac{1}{\Gamma(\gamma n)} e^{-\frac{nx}{a}} \left(\frac{nx}{a}\right)^{n\gamma-1} \frac{n\,dx}{a} \qquad (10.46)$$

Example 10.12. Distribution of the Mean for the Rectangular Population

The characteristic function of the distribution $dF = dx$ is

$$\int_0^1 e^{itx} dx = \frac{e^{it} - 1}{it}$$

The c.f. of the mean of n values is then $\left(\frac{e^{\frac{it}{n}} - 1}{\frac{it}{n}}\right)^n$, and the frequency function is thus

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \left(\frac{e^{\frac{it}{n}} - 1}{\frac{it}{n}}\right)^n dt. \qquad (10.47)$$

This integral is everywhere holomorphic and the range of integration may then be changed to the contour Γ consisting of the real axis from $-\infty$ to $-c$, the small semicircle of radius c and centre at the origin, and the real axis from c to ∞. Thus

$$f(x) = \frac{1}{2\pi} \int_{\Gamma} e^{-itx} \left(\frac{e^{\frac{it}{n}} - 1}{\frac{it}{n}}\right)^n dt$$

$$= \frac{(-1)^n}{2\pi} \int_{\Gamma} e^{-itx} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \frac{e^{\frac{itj}{n}}}{\left(\frac{it}{n}\right)^n} dt. \qquad (10.48)$$

Now

$$\int_{\Gamma} \frac{e^{igz}}{z^n} dz = 0 \text{ if } g > 0$$

$$= -2\pi i^n \frac{g^{n-1}}{(n-1)!} \text{ if } g \leqslant 0.$$

This may be seen by integrating along a contour consisting of Γ and the infinite semicircle *above* the real axis if $g > 0$ and *below* it if $g \leqslant 0$.

Substituting in (10.48) we find

$$f(x) = \frac{(-1)^n}{2\pi}\int_\Gamma \Sigma(-1)^j\binom{n}{j}\frac{e^{\frac{ij}{n}t - itx}}{\left(\frac{it}{n}\right)^n}dt$$

$$= \frac{(-1)^{n-1}}{(n-1)!}\sum_{j<nx}(-1)^j\binom{n}{j}\frac{\left(\frac{j}{n}-x\right)^{n-1}}{\left(\frac{1}{n}\right)^n}$$

$$= \frac{n^n}{(n-1)!}\sum_{j<nx}(-1)^j\binom{n}{j}\left(x-\frac{j}{n}\right)^{n-1}$$

This, with a few changes of notation, is the same as (10.40).

10.6. General expressions may also be derived for the distributions of geometric means and the moments about fixed points.

In fact, if $y = \log x$, the characteristic function of y is

$$\alpha(t) = \int_{-\infty}^{\infty} e^{it\log x}\,dF = \int_{-\infty}^{\infty} x^{it}\,dF.$$

The distribution of the sum of n independent values of y, say nz, is then given by

$$F(nz) - F(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1-e^{-itnz}}{it}\alpha \quad . \quad (10.49)$$

and the distribution of the mean is that of z. But $z = \log u$, where u is the geometric mean, and hence the distribution of u may be found.

The frequency function, when it exists, is

$$f(nz) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itnz}\,\alpha^n\,dt.$$

Similarly the characteristic function of a power of the variate, say x^r, is given by

$$\beta(t) = \int_{-\infty}^{\infty} e^{itx^r}\,dF$$

and thus the distribution of the rth moment, say z, by

$$F(nz) - F(0) = \int_{-\infty}^{\infty}\frac{e^{-itnz}-1}{it}\beta^n\,dt. \quad . \quad (10.50)$$

Example 10.13. Distribution of the Geometric Mean in Samples from a Rectangular Population

If the population is

$$dF = \frac{1}{a}dx \qquad 0 \leqslant x \leqslant a,$$

the characteristic function of $\log x$ is

$$\int_0^a x^{it}\frac{dx}{a} = \frac{a^{it}}{1+it}.$$

The *frequency* function of $u = \Sigma \log x$ is then given by

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itu}\,a^{nit}}{(1+it)^n}dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it(n\log a - u)}}{(1+it)^n}dt \qquad n\log a - u \geqslant 0.$$

This integral may be evaluated in the manner of Example 10.12 and we find

$$f(u) = \frac{(n\log a - u)^{n-1}\,e^{-(n\log a - u)}}{\Gamma(n)},$$

whence, putting $z = e^{\frac{u}{n}}$, we find for the distribution of the geometric mean z

$$f(z) = \frac{n^n z^{n-1}}{a^n\Gamma(n)}\left(\log\frac{a}{z}\right)^{n-1} \qquad . \quad . \quad . \quad . \quad . \quad (10.51)$$

Example 10.14. Distribution of the Second-order Moment about the Population Mean in Samples from a Normal Population

If the distribution is

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{x^2}{2\sigma^2}}\,dx$$

the characteristic function of x^2 is

$$\frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{itx^2}\,e^{-\frac{x^2}{2\sigma^2}}\,dx$$

$$= \frac{1}{(1-2\sigma^2 it)^{\frac{1}{2}}}.$$

The c.f. of the mean of n values, say m_2, is then

$$\frac{1}{\left(1-\frac{2\sigma^2 it}{n}\right)^{\frac{n}{2}}} \qquad . \quad (10.52)$$

and the frequency function of this is

$$\int_{-\infty}^{\infty}\frac{e^{-itm_2}}{\left(1-\frac{2\sigma^2 it}{n}\right)^{\frac{n}{2}}}\,dt.$$

This may be integrated in the manner of the previous example, or the result written down directly from the consideration that (10.52) is the characteristic function of the distribution

$$dF = \frac{n^{\frac{n}{2}}}{(2\sigma^2)^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\,e^{-\frac{nm_2}{2\sigma^2}}\,m_2^{\frac{n}{2}-1}\,dm_2, \qquad . \quad (10.53)$$

a result which may be compared with that of Example (10.5), to which it is equivalent.

The Method of Induction

10.7. The distribution of the sum of two independent variates may be obtained directly without the intervention of characteristic functions. If $F_1(x_1)$ and $F_2(x_2)$ are the distribution functions, the distribution function of $z = x_1 + x_2$ is given by

$$dF = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x_2} dF_1\,dF_2, \quad . \quad . \quad . \quad (10.54)$$

the domain of integration being that for which $x_1 + x_2 \leqslant z$

$$= \int_{-\infty}^{\infty} \Big[F_1 \Big]_{-\infty}^{z-x_2} dF_2$$

$$= \int_{-\infty}^{\infty} F_1(z - x_2)\, dF_2. \qquad . \quad . \quad . \qquad (10.55)$$

If, further, F is differentiable, the frequency function of z is given by

$$f(z) = \int_{-\infty}^{\infty} f_1(z - x_2)\, f_2\, dx_2, \qquad . \quad . \quad . \qquad (10.56)$$

f_1 and f_2 being the frequency functions of x_1 and x_2.

(10.56) can be used to obtain successively the distribution of the sum of any number of variables whose individual distributions are known. If all the variables have the same distribution the general form may be suggested when the results for two or three variates have been worked out. Its correctness can then be verified by induction. The following examples illustrate the method.

Example 10.15

Consider again the distribution

$$dF = \frac{dx}{\pi(1 + x^2)} \qquad - \infty \leqslant x \leqslant \infty.$$

By (10.56) the distribution of the sum of two independent variables each of which has this distribution has the frequency function

$$\int_{-\infty}^{\infty} \frac{1}{\pi^2}\left(\frac{1}{1 + (z - x)^2}\right)\left(\frac{1}{1 + x^2}\right) dx = \frac{2}{\pi(z^2 + 2^2)}.$$

This suggests the general form

$$\frac{n}{\pi(z^2 + n^2)}.$$

If this is correct, then the form for $(n + 1)$ variables is

$$\int_{-\infty}^{\infty} \frac{n}{\pi^2}\left(\frac{1}{n^2 + (z - x)^2}\right)\left(\frac{1}{1 + x^2}\right) dx$$

$$= \frac{n}{\pi}\left(\frac{n + 1}{n}\right)\Big/ \{z^2 + (n + 1)^2\}$$

$$= \frac{(n + 1)}{\pi\{z^2 + (n + 1)^2\}}.$$

The result holds for $n = 1$, 2, and is therefore true in general.

Example 10.16

In Example 10.4 we found that the distribution of the sums of squares of n independent variates is given by

$$dF = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-2)}\, dz. \qquad (10.57)$$

Suppose we had surmised this form from an examination of a few cases for low n. Let x be another variate distributed normally about zero mean with unit variance. We require the distribution of $z + x^2$.

Let $x^2 = v$. Then v has the distribution

$$dF = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{1}{2}v} v^{-\frac{1}{2}} \, dv.$$

Then, from (10.56) the frequency function of the distribution of $u = z + v$ is given by

$$\int_{-\infty}^{\infty} \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{1}{2}(u-z)} (u-z)^{-\frac{1}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-2)} \, dz$$

$$= \frac{e^{-\frac{1}{2}u}}{2^{\frac{n+3}{2}}\Gamma(\frac{1}{2})\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{\infty} (u-z)^{-\frac{1}{2}} z^{\frac{1}{2}(n-2)} \, dz$$

$$= \frac{B(\frac{1}{2}, \frac{1}{2}n)}{2^{\frac{n+1}{2}}\Gamma(\frac{1}{2})\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}u} u^{\frac{1}{2}(n-1)}$$

$$= \frac{e^{-\frac{1}{2}u} u^{\frac{1}{2}(n-1)}}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)}$$

which is the same as (10.57) with $n + 1$ for n. Hence the distribution holds generally.

The Distribution of a Ratio

10.8. Cases not infrequently arise in which we wish to find the sampling distribution of the ratio of two statistics, z_1, z_2. The problem becomes somewhat complicated when the divisor z_2 may be negative, but relatively simple in the contrary case.

If F_1, F_2 are the distribution functions of z_1 and z_2 and $v = \frac{z_1}{z_2}$, then for the distribution function of v we have

$$dF = \int_{-\infty}^{\infty} \int_{-\infty}^{vz_2} dF_1 \, dF_2 \, dv$$

$$= \int_{-\infty}^{\infty} F_1(vz_2) \, dF_2 \, dv, \qquad (10.58)$$

or, in terms of frequency functions,

$$f(v) = \int_{-\infty}^{\infty} z_2 f_1(vz_2) f_2(z_2) \, dz_2. \qquad (10.59)$$

Example 10.17

Consider again the distribution of the ratio $\bar{x}/s$ discussed in Example 10.6. Here $\bar{x}$ is the mean of samples of n from a normal population and is thus distributed as

$$dF \propto e^{-\frac{n\bar{x}^2}{2\sigma^2}} \, d\bar{x}.$$

s is distributed as

$$dF \propto e^{-\frac{ns^2}{2\sigma^2}} s^{n-2} \, ds,$$

as we have found in equation (10.29).

Then the distribution of $v = \dfrac{\bar{x}}{s}$ is, from (10.59), a constant time

$$\int_{-\infty}^{\infty} s \,.\, e^{-\frac{ns^2v^2}{2\sigma^2}}\, e^{-\frac{ns^2}{2\sigma^2}}\, s^{n-2}\, ds \propto \frac{1}{(1+v^2)^{\frac{n}{2}}}\,.$$

which then gives us the distribution (10.30) on the evaluation of the constant.

Example 10.18. Fisher's z-distribution

Suppose we have two independent samples of n_1 and n_2 members respectively from normal populations with variances σ_1^2 and σ_2^2. The distributions of the sample variances s_1^2 and s_2^2 $\left(= \frac{1}{n}\Sigma(x-\bar{x})^2\right)$ are then

$$dF \propto e^{-\frac{n_1 s_1^2}{2\sigma_1^2}}\, s_1^{n_1-2}\, ds_1 \qquad 0 \leqslant s_1 \leqslant \infty$$

$$dF \propto e^{-\frac{n_2 s_2^2}{2\sigma_2^2}}\, s_2^{n_2-2}\, ds_2 \qquad 0 \leqslant s_2 \leqslant \infty$$

The distribution of the ratio $t = \dfrac{s_1}{s_2}$ is then, from (10.59), given by

$$\begin{aligned} f(t) &\propto \int_0^{\infty} s_2 \exp\left(-\frac{n_1t^2s_2^2}{2\sigma_1^2}\right)(s_2t)^{n_1-2}\exp\left(-\frac{n_2s_2^2}{2\sigma_2^2}\right)s_2^{n_2-2}\, ds_2 \\ &\propto \int_0^{\infty} \exp\left\{\left(-\frac{n_1t^2}{2\sigma_1^2}-\frac{n_2}{2\sigma_2^2}\right)s_2^2\right\} s_2^{n_1+n_2-3}\, t^{n_1-2}\, ds_2 \\ &\propto \frac{t^{n_1-2}}{\left(\frac{n_1t^2}{\sigma_1^2}+\frac{n_2}{\sigma_2^2}\right)^{\frac{n_1+n_2-2}{2}}}, \qquad 0 \leqslant t \leqslant \infty \quad . \end{aligned} \tag{10.60}$$

This is usually expressed in a somewhat different form. Put

$$z = \tfrac{1}{2}\log\frac{n_1(n_2-1)s_1^2}{n_2(n_1-1)s_2^2} = \tfrac{1}{2}\log\frac{n_1(n_2-1)}{n_2(n_1-1)}t^2.$$

We find for the frequency function of z

$$f(z) \propto \frac{e^{(n_1-1)z}}{\left(\frac{(n_1-1)e^{2z}}{\sigma_1^2}+\frac{n_2-1}{\sigma_2^2}\right)^{\frac{1}{2}(n_1+n_2-2)}} \qquad -\infty \leqslant z \leqslant \infty$$

or, writing $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$, and evaluating the constant term

$$f(z) = \frac{\sigma_1^{\nu_1}\sigma_2^{\nu_2}\,\nu_1^{\frac{1}{2}\nu_1}\nu_2^{\frac{1}{2}\nu_2}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}\cdot\frac{e^{\nu_1 z}}{\left(\frac{\nu_1e^{2z}}{\sigma_1^2}+\frac{\nu_2}{\sigma_2^2}\right)^{\frac{1}{2}(\nu_1+\nu_2)}} \tag{10.61}$$

In particular, if $\sigma_1^2 = \sigma_2^2$ we get Fisher's z-distribution of the ratio of two variances from a normal population

$$f(z) = \frac{\nu_1^{\frac{1}{2}\nu_1}\nu_2^{\frac{1}{2}\nu_2}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}\,\frac{e^{\nu_1 z}}{(\nu_1e^{2z}+\nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}. \tag{10.62}$$

The distribution function of z may be obtained from tables of the incomplete B-function. Special tables showing, for various values of ν_1 and ν_2, the values of z corresponding to $F(z) = 0.99$ and 0.95, have been prepared and are given as Appendix Tables 4 and 5.

10.9. Up to this point we have been mainly concerned with the distribution of a single statistic compiled from the members of a sample which is random and simple. The methods may, however, readily be generalized to obtain the simultaneous distribution of several statistics. For example, if there are several statistics $z_1, z_2 \ldots z_p$, and the joint distribution of the sample values $x_1 \ldots x_n$ is represented by $dF(x_1, \ldots x_n)$, the characteristic function of the z's is given by

$$\phi(t_1, \ldots t_p) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp(it_1z_1 + \ldots + it_pz_p)\, dF(x_1 \ldots x_n) \qquad (10.63)$$

and the frequency function of the z's (if it exists) by

$$f(z_1, \ldots z_p) = \frac{1}{(2\pi)^p}\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp(-it_1z_1 \ldots - it_pz_p)\phi(t_1 \ldots t_n)\, dt_1 \ldots dt_p \qquad (10.64)$$

Examples of the use of these results will occur in the sequel.

NOTES AND REFERENCES

A systematic account of the various methods for deriving sampling distributions has not previously been given, except in regard to characteristic functions, as to which see Kullback (1934). The geometrical method is largely due to R. A. Fisher, whose use of it to derive the sampling distribution of the correlation coefficient (1915) is a beautiful example of the power of the method (cf. Chapter 14). See also Uspensky (1937).

Some of the distributions derived in the foregoing examples are classical. For "Student's" distribution see his paper of 1908 and Fisher's paper of 1925. The distribution of the sums of squares of values from a normal population was discovered by Helmert in 1876 but forgotten until Karl Pearson rediscovered it in 1900. The distribution of the mean of samples from a rectangular population is traceable as far back as Legendre (*Miscellanea Taurinensia*, 1770–73), but was forgotten and rediscovered simultaneously by Hall and Irwin (1927), the former using the geometrical method and the latter characteristic functions. For the distribution of means from Pearson curves, see Irwin (1930). For Fisher's z distribution, see his paper of 1915 and that of 1924. For the distribution of a ratio, see Cramér (1937) (Exercises 10.8–10.11 below), Geary (1930), Fieller (1932), and Nicholson (1941). The distribution of the ratio of two normal variables exhibits some unusual features; it may, for example, be bimodal.

Cramér, H. (1937), *Random Variables and Probability Distributions*, Cambridge University Press.

Fieller, E. C. (1932), "The distribution of an index in a normal bivariate population," *Biometrika*, **24**, 428.

Fisher, R. A. (1915), "The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, **10**, 507.

—— (1924), "On a distribution yielding the error functions of several well-known statistics," *Proc. International Math. Congress at Toronto*, 805.

—— (1925), "Applications of 'Student's' distribution," *Metron*, **5**, No. 3, 90.

Geary, R. C. (1930), "The frequency distribution of the quotient of two normal variables," *Jour. Roy. Statist. Soc.*, **93**, 442.

Hall, P. (1927), "The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable," *Biometrika*, **19**, 240.

Irwin, J. O. (1927), "On the frequency-distribution of the means of samples from a population having any law of frequency with finite moments," *Biometrika*, **19**, 225, and (1929), **21**, 431.

—— (1930), "On the frequency-distribution of the means of samples from populations of certain of Pearson's types," *Metron*, **7**, No. 4, 51.

Kullback, S. (1934), "An application of characteristic functions to the distribution problem of statistics," *Ann. Math. Statist.*, **5**, 263.

—— (1935), "On samples from a multivariate normal population," *Ann. Math. Statist.*, **6**, 203.

Nicholson, C. (1941), "A geometrical analysis of the frequency distribution of the ratio between two variables," *Biometrika*, **32**, 16.

Pearson, Karl (1900), "On the criterion that a given system of deviations from the probable . . . is such that it can be reasonably supposed to have arisen from random sampling," *Phil. Mag.*, **50**, 157.

"Student" (1908), "The probable error of a mean," *Biometrika*, **6**, 1.

Uspensky, J. V. (1937), *Introduction to Mathematical Probability*, McGraw-Hill, New York and London.

EXERCISES

10.1. Derive by the method of characteristic functions the expression for the sampling distribution of the mean of samples from the population

$$dF = \frac{dx}{\pi(1+x^2)}, \qquad -\infty \leqslant x \leqslant \infty.$$

10.2. Show that the distribution of the geometric mean g in samples of n from the Type III population

$$dF = \frac{e^{-x}x^{p-1}}{\Gamma(p)}\,dx \qquad 0 \leqslant x \leqslant \infty$$

is

$$dF = \frac{ng^{np-1}}{\Gamma(n)\{\Gamma(p)\}^n}\sum_{j=0}^{\infty}(-1)^{n+nj+1}\left[\frac{d^{n-1}}{dt^{n-1}}\frac{g^{nt}}{\Gamma(t-1)}\right]_{t=j}$$

(Kullback, 1934.)

10.3. Show that the difference of two values drawn at random from the Poisson population whose general term is $e^{-\lambda}\dfrac{\lambda^r}{r!}$ is distributed in the form whose general term is $e^{-2\lambda}\,T_d(2\lambda)$, where d can take all integral values from $-\infty$ to ∞ and $T_d(2\lambda)$ is Bessel's modified function of the first kind of order d and argument 2λ. (Cf. Example 4.5.)

(Irwin, 1937, *Jour. Roy. Statist. Soc.*, **100**, 415.)

10.4. Show that the distribution of the mean of samples of n from the Type II population

$$dF \propto x^{p-1}(1-x)^{p-1}\,dx \qquad p > 0,\ 0 \leqslant x \leqslant 1$$

is given by

$$f(\bar{x}) = \frac{n}{2}\pi^{\frac{n-2}{2}}\{\Gamma(p)\}^n\int_{-\infty}^{\infty}\frac{\left\{J_{p-\frac{1}{2}}\left(\frac{\beta}{2}\right)\right\}^n}{\beta^{n(p-\frac{1}{2})}}\cos(n\bar{x}\beta)\,d\beta,$$

where $J_r(z)$ is the Bessel coefficient of order r in z.

(Irwin, 1927.)

10.5. Show that the distribution of the geometric mean of n variables, one from each of the populations with frequency functions

$$\frac{x^{p-1}e^{-x}}{\Gamma(p)}, \quad \frac{x^{p+\frac{1}{n}-1}e^{-}}{\Gamma\left(p+\frac{1}{n}\right)} \qquad \frac{x^{p+\frac{n-1}{n}-1}}{\Gamma\left(p+\frac{n-1}{n}\right)}$$

is the same as the distribution of the arithmetic mean of n independent variables distributed in the first of these forms.

(Kullback, 1934.)

10.6. Show that the difference of two variates, z, each of which is distributed in the Type III form

$$dF = \frac{e^{-x}x^{p-1}}{\Gamma(p)}\,dx$$

has the frequency function

$$= \int_{-\infty}^{\infty} \frac{e^{-ixt}}{(1+t^2)^p}\,dt$$

$$f(z) = \frac{z^{\frac{2p-1}{2}}}{2^{\frac{2p-1}{2}}\Gamma(p)\Gamma(\frac{1}{2})} K_{\frac{2p-1}{2}}(z),$$

where $K_r(x)$ is the Bessel function of second order and imaginary argument.

(K. Pearson, Stouffer and David, 1932, *Biometrika*, **24**, 293.)

10.7. If a frequency function is given as the sum of a number of terms of the Type A series

$$f(x) = \alpha(x)\left\{1 + \frac{a_3}{\sigma^3}H_3\left(\frac{x}{\sigma}\right) + \ldots + \frac{a_k}{\sigma^k}H_k\left(\frac{x}{\sigma}\right)\right.$$

show that the sum S of n independent variates has a frequency function

$$f(S) = \alpha(S)\left\{1 + \frac{A_3}{\Sigma^3}H_3\left(\frac{S}{\Sigma}\right) + \ldots \frac{A_k}{\Sigma^k}H_k\left(\frac{S}{\Sigma}\right)\right\}$$

where $\Sigma = \sigma\sqrt{n}$ and

$$A_j = \Sigma \frac{n!}{\nu_3!\nu_4!\ldots\nu_k!(N-\nu_3-\ldots-\nu_k)!} a_3^{\nu_3}\ldots a_k^{\nu_k},$$

the summation being taken over all values of the ν's for which

$$3\nu_3 + 4\nu_4 + \ldots + k\nu_k = j.$$

(Baker, 1930, *Ann. Math. Statist.*, **1**, 199.)

10.8. A theorem of Cramér's (1937) states that if two independent variables, x_1 and x_2, with finite mean values, distribution functions F_1 and F_2 and characteristic functions ϕ_1, ϕ_2 are such that $F_2(0) = 0$, so that x_2 is non-negative, and $\int_1^{\infty} \frac{|\phi_2(t)|}{t}\,dt$ converges, then the distribution function of $v = \frac{x_1}{x_2}$ is given by

$$F(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\phi_2(t) - \phi_1(t)\phi_2'(-tv)}{it}\,dt$$

and the frequency function, if it exists, by

$$f(v) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \phi_1(t)\phi_2'(-tv)\,dt.$$

Use this result to obtain the distributions of Examples 10.17 and 10.18.

10.9. Show that the ratio of two independent normal variables has frequency function

$$f(v) = \frac{1}{\sqrt{(2\pi)}}\,\frac{m_2\sigma_1^2 + m_1\sigma_2^2 v}{(\sigma_1^2 + \sigma_2^2 v^2)^{\frac{3}{2}}}\exp\left\{-\tfrac{1}{2}\frac{(m_1 - m_2 v)^2}{\sigma_1^2 + \sigma_2^2 v^2}\right\}$$

where m_1, σ_1 are the mean and standard deviation of the first variate, m_2, σ_2 those of the second variate, and it is assumed that m_2 is so large compared with σ_2 that the range of the second variate is effectively positive.

Hence show that $\dfrac{m_1 - m_2 v}{(\sigma_2^2 + \sigma_2^2 v^2)^{\frac{1}{2}}}$ is normally distributed about zero mean with unit variance.

(Geary, 1930.)

10.10. Show that the ratio of two independent variables distributed as

$$dF \propto e^{-\gamma_1(x-m_1)}(x - m_1)^{p_1-1}dx, \qquad 0 < m_1 \leqslant x \leqslant \infty$$
$$dF \propto e^{-\gamma_2(x-m_2)}(x - m_2)^{p_2-1}dx, \qquad 0 < m_2 \leqslant x \leqslant \infty$$

has a frequency function

$$f(v) = \frac{\gamma_1^{p_1} e^{\xi\gamma} m_2}{(p_1 - 1)!}\left\{\frac{\xi^{p_1-1}}{\left(1 + \dfrac{\gamma_1 v}{\gamma_2}\right)^{p_2}} - \binom{p_1 - 1}{1}\left(\frac{v}{\gamma_2}\right)\frac{p_2\xi^{p_1-2}}{\left(1 + \dfrac{\gamma_1 v}{\gamma_2}\right)^{p_2+1}} + \binom{p_1 - 2}{2}\left(\frac{v^2}{\gamma_2^2}\right)\frac{p_2(p_2 + 1)\xi^{p_1-3}}{\left(1 + \dfrac{\gamma_1 v}{\gamma_2}\right)^{p_2+2}} - \dots\right\}$$
$$+ \frac{\gamma_1^{p_1} e^{\xi\gamma} p_2}{(p_1 - 1)!\,\gamma_2}\left\{\frac{\xi^{p_1-1}}{\left(1 + \dfrac{\gamma_1 v}{\gamma_2}\right)^{p_2+1}} - \binom{p_1 - 1}{1}\left(\frac{v}{\gamma_2}\right)\frac{(p_2 + 1)\xi^{p_1-2}}{\left(1 + \dfrac{\gamma_1 v}{\gamma_2}\right)^{p_2+2}} + \dots\right\}$$

where $\xi = m_1 - m_2 v$. (This includes Fisher's z-distribution as a particular case.)

10.11. Show that the ratio of two variates $v = \dfrac{x_1}{x_2}$, where x_1 is distributed normally with mean m_1 and variance σ^2 and the second like a standard deviation in normal samples, i.e. with distribution function given by

$$dF \propto e^{-\gamma(s-m_2)}(s - m_2)^{p-1}\,ds \qquad 0 \leqslant m_2 \leqslant s \leqslant \infty$$

has a frequency function given by

$$f(v) = \frac{\gamma^p e^{-\frac{\xi^2}{2\sigma^2}}}{\sigma\sqrt{(2\pi)}\Gamma(p)}\left\{m_2\sum_{j=0}^{\infty}\frac{v^j\xi^j\Gamma\left(p + \dfrac{j}{2}\right)}{j!\sigma^{2j}\left(\gamma + \frac{1}{2}\dfrac{v^2}{\sigma^2}\right)^{p+\frac{j}{2}}} + \sum_{j=0}^{\infty}\frac{v^j\xi^j\Gamma\left(p + \dfrac{j+1}{2}\right)}{j!\sigma^{2j}\left(\gamma + \frac{1}{2}\dfrac{v^2}{\sigma^2}\right)^{p+\frac{j+1}{2}}}\right\}$$

where $\xi = m_1 - m_2 v$.

(This includes "Student's" distribution as a particular case.)

CHAPTER 11

APPROXIMATIONS TO SAMPLING DISTRIBUTIONS

11.1. In the previous chapter we have considered methods of deriving sampling distributions in an exact form when the parent population is completely specified. Those methods are not applicable when the parent is not completely known, and they may in any case lead to results which are difficult to apply in practice, e.g. by yielding an integral which has not been tabulated. In such cases we can frequently deal with the problem by finding approximate forms for the sampling distribution, particularly by ascertaining its lower moments and then fitting a tractable type of curve such as one of the Pearson class.

A procedure of this kind has, in fact, already been considered in Chapter 9, wherein it was seen that approximate expressions could be derived for the first and second moments of sampling distributions in terms of the lower moments of the parent. When the sampling distribution tends to normality this, in effect, solves our problem, for the first and second moments determine a normal distribution. The methods of this chapter are really developments of this idea. We shall discuss exact methods of finding the moments of sampling distributions in terms of parent moments. Our results are important not only on their own account, but in giving an accurate method of judging the degree of approximation of the expressions for large n discussed in Chapter 9. In particular we shall be able to take up some points which had to be left on one side in that chapter—e.g. the rapidity with which some functions of the moments such as $\sqrt{b_1}$ approach normality.

11.2. It is as well to recall that there are three different types of moment concerned in the investigation : (*a*) the moments of the parent population, (*b*) the moments of the sample and (*c*) the moments of the sampling distribution. They will be referred to as parent-moments (parameters), sample-moments (moment-statistics) and sampling-moments respectively. Similarly we shall consider parent-cumulants, sample-cumulants and sampling-cumulants.

11.3. In Chapter 9 we obtained the exact results

$$\left.\begin{aligned} E(m'_r) &= \mu'_r \\ \operatorname{var}(m'_r) = E(m'_r - \mu'_r)^2 &= \frac{1}{n}(\mu'_{2r} - \mu'^2_r) \end{aligned}\right\} \qquad . \quad (11.1)$$

and noted that formulae for sample moments about the mean were more difficult to obtain. Although we shall later reject this approach in favour of another, it is instructive to consider what happens if we try to generalize the procedure of that chapter to our present problem. Suppose, for example, we are interested in the sampling distribution of the variance. The above equations give us the first two sample moments of the second moment about an arbitrary point. For the first sample moment of the variance we have

$$E(m_2) = E\left[\frac{1}{n}\Sigma(x^2) - \left\{\frac{1}{n}\Sigma(x)\right\}^2\right]$$

$$= E\left[\frac{n-1}{n^2}\Sigma x^2 - \frac{1}{n^2}\Sigma x_j x_k\right] \qquad j \neq k$$

$$= \frac{n-1}{n}\mu_2 - \frac{n(n-1)}{n^2}\mu_1'^2$$

$$= \frac{n-1}{n}\mu_2. \qquad (11.2)$$

This is exact and may be compared with the approximate expression given by the methods of Chapter 9, viz.:

$$E(m_2) = \mu_2. \qquad (11.3)$$

We might then proceed to find the second, third . . . sampling moments of the variance and thus obtain more and more information about its sampling distribution. For example, we have for the fourth moment

$$E(m_2^4) = E\left[\frac{1}{n}\Sigma(x^2) - \frac{1}{n^2}\{\Sigma(x)\}^2\right]^4$$

$$= E\left[\frac{1}{n^4}\{\Sigma(x^2)\}^4 - \frac{4}{n^5}\{\Sigma(x^2)\}^3\{\Sigma(x)\} + \frac{6}{n^6}\{\Sigma(x^2)\}^2\{\Sigma(x)\}^2 - \frac{4}{n^7}\{\Sigma(x^2)\}\{\Sigma(x)\}^6 + \frac{1}{n^8}\{\Sigma(x)\}^8\right]. \qquad (11.4)$$

We can then find the expectations of the individual terms by an easy extension of the method already used. We express any power in terms of products of the type $\Sigma(x_j^\alpha x_k^\beta \ldots x_l^\rho)$ when $j \neq k \neq \ldots \neq l$; the mean value of such a product, the x's being independent, is $n(n-1)\ldots(n-l+1)\mu_\alpha'\mu_\beta' \ldots \mu_\rho'$. Without loss of generality we may take our origin at the mean of the parent, so that $\mu_1' = 0$ and other moments are those about the mean of the parent. The rest is mere algebra. For example, for the first term in (11.4) we have

$$\{\Sigma(x^2)\}^4 = (x_1^2 + x_2^2 + \ldots x_n^2)^4$$
$$= \Sigma x^8 + 4\Sigma x_j^6 x_k^2 + 6\Sigma x_j^4 x_k^2 x_l^2 + 3\Sigma x_j^4 x_k^4 + \Sigma x_j^2 x_k^2 x_l^2 x_m^2. \qquad (11.5)$$

The numerical coefficients require a little watching. That of $x_j^4 x_k^4$, for example, is 3, not 6 as in the multinomial expansion of $(x_1^2 + \ldots x_n^2)^4$ because j and k can be interchanged. The mean value of (11.5) is then

$$\mu_8 + 4n(n-1)\mu_6\mu_2 + 6n(n-1)(n-2)\mu_4\mu_2^2 + 3n(n-1)\mu_4^2 + n(n-1)(n-2)(n-3)\mu_2^4.$$

A similar evaluation of the other terms in (11.4) leads eventually to the result

$$E(m_2^4) = \frac{3}{n^2}(\mu_4 - \mu_2^2)^2 + \frac{1}{n^3}(\mu_8 - 4\mu_6\mu_2 - 24\mu_5\mu_3 - 15\mu_4^2 + 48\mu_4\mu_2^2 + 96\mu_3^2\mu_2 - 30\mu_2^4)$$
$$- \frac{1}{n^4}(4\mu_8 - 40\mu_6\mu_2 - 96\mu_5\mu_3 - 54\mu_4^2 + 336\mu_4\mu_2^2 + 528\mu_3^2\mu_2 - 306\mu_2^4)$$
$$+ \frac{1}{n^5}(6\mu_8 - 96\mu_6\mu_2 - 176\mu_5\mu_3 - 102\mu_4^2 + 924\mu_4\mu_2^2 + 1232\mu_3^2\mu_2 - 1044\mu_2^4)$$
$$- \frac{1}{n^6}(4\mu_8 - 88\mu_6\mu_2 - 160\mu_5\mu_3 - 95\mu_4^2 + 1050\mu_4\mu_2^2 + 1360\mu_3^2\mu_2 - 1395\mu_2^4)$$
$$+ \frac{1}{n^7}(\mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 + 560\mu_3^2\mu_2 - 630\mu_2^4) \qquad (11.6)$$

11.4. Systematic investigations of the sampling moments on these lines (though by

a somewhat different method) were carried out by Tschuprow (1919) and, for the particular case of the variance, by Church (1925), who corrected some misprints in Tschuprow's results. Unfortunately the resulting formulae are exceedingly complicated—the above is one of the simpler cases—and are obviously unsuitable for practical work.

It then began to be appreciated that their complexity might be due to the use of a special type of symmetric function of the observations, namely the moments, and the question arose whether other functions might have simpler properties. Thiele had already introduced the parameters which are now known as cumulants, and had defined some statistics which were the same functions of the moment-statistics as the cumulants are of the moments. He also gave some expressions for the sampling cumulants of these functions. In 1928 C. C. Craig developed this work and gave a number of further results. Even these, however, were sufficiently complicated and were reached only after some labour, and Craig himself remarked that "it rather seems that the best hopes of effectively further simplifying the problem of sampling for statistical characteristics lie either in the discovery of a new kind of symmetric functions of all the observations . . . or in the abandonment of the method of characterizing frequency functions by symmetric functions of the observations altogether."

About the same time R. A. Fisher discovered such a new kind of symmetric function, the k-statistics, and his remarkable paper of 1928 forms the basis of nearly all subsequent work on the subject. The new statistics have the valuable property of yielding particularly simple sampling formulae which can be obtained directly by combinatorial methods, obviating most of the algebraic labour inherent in the older methods.

Seminvariant Statistics

11.5. It will be observed that equation (11.6) does not contain the parent-mean μ_1'. In deriving it we took an arbitrary mean at the parent mean, which simplified the algebra to some extent. The independence of $E(m_2^4)$ of this parent mean is, however, not due to this accidental circumstance. In fact any transformation of the variate from one origin to another leaves m_2 unchanged, for $m_2 = \Sigma(x - m_1')^2$ and the transformation increases each x and m_1' by the same amount, leaving their difference unaffected. Consequently if m_2 is independent of the location of the origin, so must be its sampling moments. Thus our sampling formulae are very much simplified if we use statistics which are independent of the origin. In equation (11.6) there are terms corresponding to μ_8, $\mu_6\mu_2$, μ_4^2, $\mu_4\mu_2^2$ and μ_2^4. If we had to take account of possible terms in μ_1' there would be additional terms such as $\mu_7'\mu_1'$, $\mu_6'\mu_1'^2$, and so on, our formula containing 22 types of term instead of only 5.

11.6. A statistic which is independent of the origin of calculation is said to be seminvariant. The moment-statistics about the mean are seminvariant. We now consider a second family of statistics k_p $(p = 1, 2, \ldots)$, symmetric in the observations $x_1 \ldots x_n$, such that the mean value of k_p is the pth cumulant, i.e.

$$E(k_p) = \kappa_p. \qquad \qquad (11.7)$$

Note first of all that k_p is uniquely determined by this definition; for if there were two functions k_p and k_p' obeying (11.7) their difference $k_p - k_p'$ would have a zero mean value. But this difference is itself a symmetric function and can therefore be expressed as the sum of terms Σx^p, $\Sigma x_j x_k^{p-1}$, etc., and hence its mean value is a series of terms each of which is a product of moments. The vanishing of this series would imply a relationship among the moments which is impossible except perhaps for particular parent populations. Hence $k_p - k_p'$ must vanish identically and thus $k_p = k_p'$.

Secondly, note that the k's are in fact seminvariant, except for k_1 which is equal to the mean itself. In fact, we have by Taylor's theorem

$$k_p(x_1+h, x_2+h, \ldots x_n+h) = k_p(x_1, x_2, \ldots x_n) + \frac{h}{1!}Dk_p(x_1, x_2, \ldots x_n) + \frac{h^2}{2!}D^2k_p(x_1, x_2, \ldots x_n) + \ldots \quad . \quad (11.8)$$

where

$$D = \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2} + \qquad \frac{\partial}{\partial x_n}$$

Taking mean values, and remembering that κ_p itself is independent of the origin, except for κ_1, we have

$$\kappa_p = \kappa_p + \frac{h}{1!}E(Dk_p) +, \text{etc.} \quad . \quad . \quad . \quad . \quad (11.9)$$

Thus $E(Dk_p)$ and other terms on the right vanish separately, for (11.9) is an identity in h. In virtue of the remark above, this implies that $Dk_p = 0$, $D^2k_p = 0$, and so on; and hence, from (11.8),

$$k_p(x_1+h, x_2+h, \ldots x_n+h) = k_p(x_1, x_2, \ldots x_n).$$

i.e. k_p is seminvariant. The exception to this rule is k_1 which has as its mean value $\kappa_1 = m_1$ and thus

$$k_1 = \frac{1}{n}\Sigma(x). \quad . \quad (11.10)$$

11.7. We now proceed to find explicit expressions for the k-statistics in terms of the observations $x_1 \ldots x_n$. By definition k_p is degree p in these observations (for κ_p is of order p in the moments, that is, the sum of the orders of the moments comprising any term in κ_p is p). We may then write

$$k_p = \Sigma\Sigma(x_1^{p_1}x_2^{p_1} \ldots x_{\pi_1}^{p_1}x_{\pi_1+1}^{p_2} \ldots x_{\pi_1+\pi_2}^{p_2} \ldots x_{\pi_1+\ldots+\pi_s}^{p_s})A(p_1^{\pi_1} \ldots p_s^{\pi_s}) \quad (11.11)$$

where the second summation extends over all the ways of assigning the $\pi_1 + \pi_2 + \ldots \pi_s$ subscripts (including permutations) from the n available and the first summation extends over all partitions of the number p, $(p_1^{\pi_1}p_2^{\pi_2} \ldots p_s^{\pi_s})$. $A(p_1^{\pi_1} \ldots p_s^{\pi_s})$ is a number depending on the partition.

We have

$$p_1\pi_1 + p_2\pi_2 + \ldots + p_s\pi_s = p \quad . \quad . \quad . \quad . \quad (11.12)$$

and define ρ by

$$\pi_1 + \pi_2 + \ldots + \pi_s = \rho. \quad . \quad . \quad . \quad . \quad (11.13)$$

On taking mean values of (11.11) we have, since the x's are independent,

$$\kappa_p = \Sigma\{(\mu_{p_1}^{\pi_1}\mu_{p_2}^{\pi_2} \ldots \mu_{p_s}^{\pi_s})AB\}. \quad . \quad . \quad . \quad . \quad (11.14)$$

where B is the number of ways of picking out the ρ subscripts from n, permutations allowed, and is therefore equal to $n(n-1) \ldots (n-\rho+1) \quad n^{(\rho)}$

Now from equation (3.31), we have

$$\kappa_p = p!\Sigma\Sigma\left(\frac{\mu_{p_1}}{p_1!}\right)^{\pi_1} \ldots \left(\frac{\mu_{p_s}}{p_s!}\right)^{\pi_s}\frac{(-1)^{\rho-1}(\rho-1)!}{\pi_1! \ldots \pi_s!}. \quad . \quad (11.15)$$

the summation extending over all partitions subject to (11.12) and (11.13). On identifying corresponding terms in (11.14) and (11.15) we find the values of the A's and on substituting in (11.11) obtain finally

$$k_p = \frac{p!\Sigma(-1)^{\rho-1}(\rho-1)!}{n^{[\rho]}} \Sigma \frac{x_1^{p_1} \ldots x_\rho^{p_s}}{(p_1!)^{\pi_1} \ldots (p_s!)^{\pi_s}\pi_1! \ldots \pi_s!}, \quad \ldots \quad (11.16)$$

the explicit expression of k_p in terms of the x's.

We may notice an important simplification of this expression which is crucial in a discussion of the sampling properties of the k's. Apart from factors in ρ and n a typical term in (11.16) may be written

$$p!\left(\frac{x_1^{p_1}}{p_1!}\frac{x_2^{p_1}}{p_1!} \ldots \frac{x_{\pi_1}^{p_1}}{p_1!} \ldots \frac{x_\rho^{p_s}}{p_s!}\right)\cdot\frac{1}{\pi_1! \ldots \pi_s!}$$

where, it is to be remembered, permutations of the subscripts are allowed. There will be a term of this type corresponding to every partition of ρ into π's and of p into p's. Consequently we may write

$$k_p = \Sigma\frac{(-1)^{\rho-1}(\rho-1)!}{n^{[\rho]}}\Sigma(x_{\gamma_1}x_{\gamma_2} \ldots x_{\gamma_p}), \quad \ldots \quad (11.17)$$

where there is a term in the second summation corresponding to every possible way of assigning the subscripts. In this assignment subscripts are regarded as distinct entities. For example, if from the n subscripts we choose p_1 to be 1, p_1 to be 2, ... p_2 to be $\pi_1 + 1$, and so on, there will be as many different terms as there are ways of choosing p_1 from the 1's, and so on, i.e.

$$\frac{p!}{(p_1!)^{\pi_1} \ldots (p_s)^{\pi_s}\pi_1! \ldots \pi_s!}. \quad \ldots \quad (11.18)$$

In fact, (11.16) is a condensed form of (11.17) in which all the terms leading to the same x-product are added together, their number being given by (11.18).

Expression of k-Statistics in terms of Symmetric Products and Sums

11.8. Writing

$$[p_1^{\pi_1} p_2^{\pi_2} \ldots p_s^{\pi_s}] = \Sigma(x_i^{p_1} x_j^{p_1} \ldots x_l^{p_s}) \qquad i \neq j \neq \ldots \neq l \quad . \quad (11.19)$$

so that, for instance,

$$[21] = \Sigma(x_i^2 x_j)$$

$$[2^21] = \Sigma(x_i^2 x_j^2 x_k)$$

we see that the mean value of $[p_1^{\pi_1} \ldots p_s^{\pi_s}]$ is $n^{[\rho]}\mu_{p_1}^{\pi_1} \ldots \mu_{p_s}^{\pi_s}$. We can then write down the k's in terms of the symmetric product sums $[p^\pi]$ at once from the expressions of cumulants in terms of moments. For instance, from (3.33) we have $\kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$ and hence

$$k_3 = \frac{[3]}{n} - \frac{3[21]}{n^{[2]}} + \frac{2[1^3]}{n^{[3]}}$$

$$= \frac{[3]}{n} - \frac{3[21]}{n(n-1)} + \frac{2[1^3]}{n(n-1)(n-2)} \quad \ldots \quad (11.20)$$

a result which, of course, can be obtained directly from (11.16). In fact, there are three partitions of 3, (3), (21), and (1^3). From (11.16) we then have

$$k_3 = \frac{1}{n}.\frac{3![3]}{(3!)^1 1!} + \frac{(-1)1!3![21]}{n(n-1)2!1!1!1!} + \frac{(-1)^2 2!3![1^3]}{n(n-1)(n-2)(1!)^3 3!}$$

$$= \frac{[3]}{n} - \frac{3[21]}{n(n-1)} + \frac{2[1^3]}{n(n-1)(n-2)}$$

as before.

It is, however, more useful for practical calculation of the k-statistics to express them in terms of the power sums defined by

$$s_r = \Sigma(x^r). \qquad . \quad . \quad . \quad . \quad . \quad . \quad (11.21)$$

This can be done by expressing the product sums (11.19) in terms of power sums (a procedure which may be facilitated by the use of tables of symmetric functions) or directly as follows :—

Assume

$$k_3 = a_0 s_3 + a_1 s_2 s_1 + a_2 s_1^3.$$

Since $E(k_3) = \kappa_3 = \mu_3$ we have

$$\mu_3 = a_0 E(s_3) + a_1 E(s_1 s_2) + a_2 E(s_1^3).$$

Hence, for moments about an arbitrary point

$$\begin{aligned} \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = {} & a_0(n\mu_3') + a_1\{n\mu_3' + n(n-1)\mu_2'\mu_1'\} \\ & + a_2\{n\mu_3' + 3n(n-1)\mu_2'\mu_1' + n(n-1)(n-2)\mu_1'^3\}, \end{aligned}$$

from which we find, identifying coefficients,

$$\begin{aligned} 1 &= n(a_0 + a_1 + a_2) \\ -3 &= n(n-1)(a_1 + 3a_2) \\ 2 &= n(n-1)(n-2)a_2 \end{aligned}$$

whence, solving for a_0, a_1 and a_2, we find

$$k_3 = \frac{1}{n^{[3]}}(n^2 s_3 - 3n s_2 s_1 + 2s_1^3).$$

11.9. The first eight k-statistics in terms of the power sums are as follows :—

$$\begin{aligned} k_1 &= \frac{1}{n}s_1 \\ k_2 &= \frac{1}{n^{[2]}}(ns_2 - s_1^2) \\ k_3 &= \frac{1}{n^{[3]}}(n^2 s_3 - 3ns_2 s_1 + 2s_1^3) \qquad\qquad (11.22) \\ k_4 &= \frac{1}{n^{[4]}}\{(n^3 + n^2)s_4 - 4(n^2 + n)s_3 s_1 - 3(n^2 - n)s_2^2 + 12ns_2 s_1^2 - 6s_1^4\} \\ k_5 &= \frac{1}{n^{[5]}}\{(n^4 + 5n^3)s_5 - 5(n^3 + 5n^2)s_4 s_1 - 10(n^3 - n^2)s_3 s_2 + 20(n^2 - 2n)s_3 s_1^2 \\ &\quad + 30(n^2 - n)s_2^2 s_1 - 60ns_2 s_1^3 + 24s_1^5\} \end{aligned}$$

$$
\begin{aligned}
k_6 = \frac{1}{n^{[6]}}\{&(n^5 + 16n^4 + 11n^3 - 4n^2)s_6 - 6(n^4 + 16n^3 + 11n^2 - 4n)s_5s_1 \\
&- 15(n^4 - 4n^3 - n^2 + 4n)s_4s_2 - 10(n^4 - 2n^3 + 5n^2 - 4n)s_3^2 \\
&+ 30(n^3 + 9n^2 + 2n)s_4s_1^2 + 120(n^3 - n)s_3s_2s_1 + 30(n^3 - 3n^2 + 2n)s_2^3 \\
&- 120(n^2 + 3n)s_3s_1^3 - 270(n^2 - n)s_2^2s_1^2 + 360ns_2s_1^4 - 120s_1^6\}
\end{aligned}
$$

$$
\begin{aligned}
k_7 = \frac{1}{n^{[7]}}\{&(n^6 + 42n^5 + 119n^4 - 42n^3)s_7 - 7(n^5 + 42n^4 + 119n^3 - 42n^2)s_6s_1 \\
&- 21(n^5 + 12n^4 - 31n^3 + 18n^2)s_5s_2 - 35(n^5 + 5n^3 - 6n^2)s_4s_3 \\
&+ 42(n^4 + 27n^3 + 44n^2 - 12n)s_5s_1^2 + 210(n^4 + 6n^3 - 13n^2 + 6n)s_4s_2s_1 \\
&+ 140(n^4 + 5n^2 - 6n)s_3^2s_1 + 210(n^4 - 3n^3 + 2n^2)s_3s_2^2 \\
&- 210(n^3 + 13n^2 + 6n)s_4s_1^3 - 1260(n^3 + n^2 - 2n)s_3s_2s_1^2 \\
&- 630(n^3 - 3n^2 + 2n)s_2^3s_1 + 840(n^2 + 4n)s_3s_1^4 + 2520(n^2 - n)s_2^2s_1^3 \\
&- 2520ns_2s_1^5 + 720s_1^7\}
\end{aligned}
\tag{11.22}
$$

$$
\begin{aligned}
k_8 = \frac{1}{n^{[8]}}\{&(n^7 + 99n^6 + 757n^5 + 141n^4 - 398n^3 + 120n^2)s_8 - 8(n^6 + 99n^5 + 757n^4 \\
&+ 141n^3 - 398n^2 + 120n)s_7s_1 - 28(n^6 + 37n^5 - 39n^4 - 157n^3 \\
&+ 278n^2 - 120n)s_6s_2 - 56(n^6 + 9n^5 - 23n^4 + 111n^3 - 218n^2 + 120n)s_5s_3 \\
&- 35(n^6 + n^5 + 33n^4 - 121n^3 + 206n^2 - 120n)s_4^2 + 56(n^5 + 68n^4 + 359n^3 \\
&- 8n^2 - 60n)s_6s_1^2 + 336(n^5 + 23n^4 - 31n^3 - 23n^2 + 30n)s_5s_2s_1 \\
&+ 560(n^5 + 5n^4 + 5n^3 + 5n^2 - 6n)s_4s_3s_1 + 420(n^5 + 2n^4 - 25n^3 \\
&+ 46n^2 - 24n)s_4s_2^2 + 560(n^5 - 4n^4 + 11n^3 - 20n^2 + 12n)s_3^2s_2 \\
&- 336(n^4 + 38n^3 + 99n^2 - 18n)s_5s_1^3 - 2520(n^4 + 10n^3 - 17n^2 + 6n)s_4s_2s_1^2 \\
&- 1680(n^4 + 2n^3 + 7n^2 - 10n)s_3^2s_1^2 - 5040(n^4 - 2n^3 - n^2 + 2n)s_3s_2^2s_1 \\
&- 630(n^4 - 6n^3 + 11n^2 - 6n)s_2^4 + 1680(n^3 + 17n^2 + 12n)s_4s_1^4 \\
&+ 13{,}440(n^3 + 2n^2 - 3n)s_3s_2s_1^3 + 10{,}080(n^3 - 3n^2 + 2n)s_2^3s_1^2 \\
&- 6720(n^2 + 5n)s_3s_1^5 - 25{,}200(n^2 - n)s_2^2s_1^4 + 20{,}160ns_2s_1^6 - 5040s_1^8\}
\end{aligned}
$$

In particular, we have

$$
\left.
\begin{aligned}
k_1 &= m_1 \\
k_2 &= \frac{n}{n-1}m_2 \\
k_3 &= \frac{n^2}{(n-1)(n-2)}m_3 \\
k_4 &= \frac{n^2}{(n-1)(n-2)(n-3)}\{(n+1)m_4 - 3(n-1)m_2^2\}
\end{aligned}
\right\}
\tag{11.23}
$$

expressing the k's in terms of the moment statistics.

11.10. There is a well-known theorem of symmetric functions which states that any rational integral algebraic symmetric function of $x_1 \ldots x_n$ can be expressed uniquely, rationally, integrally and algebraically in terms of the symmetric sums s_r. It can thus be so expressed in terms of the k's, for from equations such as (11.22) the s's can be so expressed in terms of the k's. Thus an investigation of the sampling constants of any symmetric function expressible in terms of rational integral algebraic symmetric functions can be translated into an investigation concerning the k's.

To round off this account of the relationship between the k's and the s's we may refer to two interesting operational properties. Write K_p for the same function of the differential

operators $\frac{\partial}{\partial x_1} \ldots \frac{\partial}{\partial x}$ as k_p is of the x's and S_p for the same function of the operators as s_p is of the x's. Then

$$\left.\begin{aligned} K_p s_p &= p! \\ K_p(s_{p_1} s_{p_2} \ldots s_{p_m}) &= 0 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (11.24)$$

where $(p_1 \ldots p_m)$ is any partition of p other than p itself: and

$$\left.\begin{aligned} S_p k_p &= p! \\ S_q k_p &= 0, \qquad q \neq p \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (11.25)$$

Methods of proof and applications of these results are given in the exercises at the end of the chapter.

Sampling Cumulants of k-Statistics

11.11. The problem of determining the sampling moments or the sampling cumulants of k-statistics is that of finding mean values of powers and products of those statistics. To any number a with partition $(a_1^{\alpha_1} a_2^{\alpha_2} \ldots a_s^{\alpha_s})$ there will correspond a moment

$$\mu(a_1^{\alpha_1} \ldots a_s^{\alpha_s}) = E(k_{a_1}^{\alpha_1} \ldots k_{a_s}^{\alpha_s}) \quad . \quad . \quad . \quad . \quad (11.26)$$

and a cumulant $\kappa\,(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ related to the moments by the identity (cf. equation (3.54))

$$\Sigma\left\{\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})\frac{t_{a_1}^{\alpha_1}}{\alpha_1!} \ldots \frac{t_{a_s}^{\alpha_s}}{\alpha_s!}\right\} = \log\left\{\Sigma\mu(b_1^{\beta_1} \ldots b_m^{\beta_m})\frac{t_{b_1}^{\beta_1}}{\beta_1!} \ldots \frac{t_{b_m}^{\beta_m}}{\beta_m!}\right\} \quad (11.27)$$

For example, the fourth cumulant of k_2 will correspond to the fourth moment of k_2, which is the mean value of k_2^4. These quantities will be written $\kappa(2^4)$ and $\mu(2^4)$, in accordance with (11.26). Again the cumulant $\kappa(32)$ corresponds to the moment $\mu(32)$, the mean value of $k_3 k_2$, or their covariance in their joint sampling distribution. Generally, in the simultaneous distribution of the k's there will be a separate formula of degree a for every partition of a.

Now the product $k_{a_1}^{\alpha_1} \ldots k_{a_s}^{\alpha_s}$ is homogeneous and of total degree a in the x's. Hence, when mean values are taken $\mu(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ will be homogeneous and of total order a in the parent μ's. Since the κ's themselves are of homogeneous order in the μ's it follows that $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ is of homogeneous order in the κ's. Hence we get the first rule for the sampling of k-statistics (which is true of seminvariants generally):—

Rule 1. $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ consists of the sum of terms each of which, except for constants, is a product of parent κ's of order a.

For instance, $\kappa(2^4)$ is of total order 8 and is therefore the sum of terms in κ_8, $\kappa_6\kappa_2$, κ_4^2, $\kappa_4\kappa_2^2$ and κ_2^4. Similarly $\kappa(32)$ will contain a term in κ_5 and one in $\kappa_3\kappa_2$ and no others. As seen in the next rule, no terms in κ_1 appear (as again is true of seminvariants generally).

Rule 2. No term in $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ contains κ_1, except $\kappa(1)$ itself.

This follows as in **11.5.** The k-statistics are seminvariant and hence their sampling distribution cannot depend on the variable quantity κ_1. The exception occurs when we are dealing with the only statistic which is dependent on the origin, namely k_1, and here $\kappa(1) = \kappa_1$ as is evident from the definitions.

11.12. We now enunciate and illustrate the rules by which the terms in $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ can be found. As the proof of the validity of the rules is difficult to grasp until their nature has been comprehended we defer a proof until later in the chapter.

To find the term in $\kappa_{b_1}{}^{\beta_1} \ldots \kappa_{b_m}{}^{\beta_m}$ in $\kappa(a_1{}^{\alpha_1} \ldots a_s{}^{\alpha_s})$ consider the two-way array

$$\begin{array}{cc|c} & & b_1 \\ & & b_1 \\ & & \\ a_1 & a_1 & \end{array} \qquad (11.28)$$

where there is a row corresponding to every κ in the term $\kappa_{b_1}{}^{\beta_1} \ldots \kappa_{b_m}{}^{\beta_m}$ and a column corresponding to every part in $\kappa(a_1{}^{\alpha_1} \ldots a_s{}^{\alpha_s})$. Consider the various ways in which the body of the table can be completed by the insertion of numbers whose row and column sums are the respective b and a numbers; e.g. if we are seeking the coefficient of $\kappa_6\kappa_2^2$ in $\kappa(4^22)$ we shall consider such arrays as

$$\begin{array}{ccc|c} 2 & 2 & 2 & 6 \\ 1 & 1 & . & 2 \\ 1 & 1 & . & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 3 & 3 & . & \\ 1 & . & 1 & 2 \\ . & 1 & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \quad . \quad (11.29)$$

Then the rules by which these arrays give the coefficients of $\kappa_{b_1}{}^{\beta_1} \ldots \kappa_{b_m}{}^{\beta_m}$ are as follows:

Rule 3. Every array in which the numbers in the body of the array fall into two or more blocks, each confined to separate rows or columns, is to be ignored.

For instance, in the foregoing example

$$\begin{array}{ccc|c} 4 & 2 & . & 6 \\ & 2 & . & 2 \\ & & 2 & 2 \\ & & & \\ 4 & 4 & 2 & 10 \end{array}$$

is to be ignored, since the 2×2 block in the top left-hand corner has no row or column number in common with the entry in the bottom right-hand corner.

Rule 4. Subject to the ignoration of terms enjoined by Rule 3, to the coefficient of $\kappa_{b_1}{}^{\beta_1} \ldots \kappa_{b_m}{}^{\beta_m}$ in $\kappa(a_1{}^{\alpha_1} \ldots a_s{}^{\alpha_s})$ there will be a contribution corresponding to each way of completing the array (11.28). Such of these as do not vanish are composed of a numerical coefficient multiplied by a function of n.

Rule 5. The numerical coefficient is the number of ways in which the column totals, considered as composed of distinct individuals, can be allocated to form the array concerned, divided by $\beta_1!\,\beta_2! \ldots \beta_m!$.

Rule 6. The function of n, called the pattern function, depends only on the configuration of zeros in the array, not on the actual numbers composing it or on the row and column totals. The function is given by considering the separations of the rows into distinct groups or separates.

(i) With one separate there is associated the number n, with two separates $n(n-1)\ldots$, with q separates $n(n-1)\ldots(n-q-1)$.

(ii) In each separation we count the number of separates in which a particular column is represented by a non-zero entry. If in ρ separates, we assign the factor

$$\frac{(-1)^{\rho-1}(\rho-1)!}{n(n-1)\ldots(n-\rho+1)}.$$

(iii) This is done for each column.

(iv) The various factors given by (ii) and (iii) are multiplied together for each separation, multiplied by the factor appropriate under (i) and the results summed to give the pattern function.

Rule 7. Any array containing a row which consists of a single non-zero entry has a vanishing pattern function and is to be ignored.

Rule 8. Any array containing a column which consists of a single non-zero entry has a pattern function $\frac{1}{n}$ times that of the array obtained by omitting that column.

Rule 9. Any array the non-zero elements of which consist of two groups connected only by a single column has a vanishing pattern function and is to be ignored.

Example 11.1

As an illustration of these rules (which are not as difficult as they look), suppose we seek for the coefficient of $\kappa_6\kappa_2^2$ in $\kappa(4^22)$. If the reader will write down the thirty or so possible arrays with column totals 4, 2, 2 and row totals 6, 2, 2, he will find that the only ones which do not vanish are those of (11.29) and permutations of rows and columns with the same sum, namely

$$\begin{array}{ccc|c} 2 & 2 & 2 & 6 \\ 1 & 1 & . & 2 \\ 1 & 1 & . & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 3 & 2 & 1 & 6 \\ 1 & 1 & . & 2 \\ . & 1 & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 1 & . & 1 & 2 \\ 1 & 1 & . & 2 \\ \hline 4 & 4 & 2 & 10 \end{array}$$

$$(a) \qquad\qquad (b) \qquad\qquad (c) \qquad\qquad (d)$$

$$\begin{array}{ccc|c} 3 & 2 & 1 & 6 \\ . & 1 & 1 & 2 \\ 1 & 1 & . & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 3 & 3 & . & 6 \\ 1 & . & 1 & 2 \\ . & 1 & 1 & 2 \\ \hline 4 & 4 & & 10 \end{array} \qquad \begin{array}{cccc} 3 & 3 & . & 6 \\ . & 1 & 1 & 2 \\ 1 & . & 1 & 2 \\ \\ 4 & 4 & 2 & 10 \end{array} \qquad . \quad . \quad (11.30)$$

$$(e) \qquad\qquad (f) \qquad\qquad (g)$$

With practice the reader will find it unnecessary to write down arrays such as (c), (d) and (e), which are merely obtained from (b) by permuting rows and columns, but for clarity at this stage they have been set out in full. There is one trap here to be particularly noticed. In array (b) the two columns summing to 4 and the two rows summing to 2 are different, and their permutations result in 4 different arrays. But in array (f), though the rows and columns are different, there are only 2 different arrays.

Each of these arrays contributes to the coefficient required. Consider first of all that from (a). The numerical coefficient is $\left(\frac{4!}{2!1!1!}\right)\left(\frac{4!}{2!1!1!}\right)\cdot\frac{1}{2!} = 72$. The first factor in brackets is the number of ways of allocating 4 individuals in the partition 2, 1, 1, similarly

for the second, and we divide by $2!$ since there are 2 members of the row totals the same, this being the only β factor.

Under Rule 8, the pattern function is $\frac{1}{n}$ times that of

$$\begin{matrix} \times & \times \\ \times & \times \\ \times & \times \end{matrix}$$

There are five separations of this, one of one separate, three of two separates and one of three separates. The contributions respectively under Rule 6 will be found to be

$$n.\left(\frac{1}{n}\right)\left(\frac{1}{n}\right) = \frac{1}{n}$$

$$3n(n-1)\left\{\frac{-1}{n(n-1)}\right\}\left\{\frac{-1}{n(n-1)}\right\} - \frac{}{n(n-1)}$$

$$n(n-1)(n-2)\ \left\{\frac{(-1)^2 2!}{n(n-1)(n-2)}\right\}\left\{\frac{(-1)^2 2!}{n(n-1)(n-2)}\right\} = \frac{4}{n(n-1)(n-2)}$$

The sum of these is $\frac{n}{(n-1)(n-2)}$ and hence the contribution from array (a) in (11.30) is $\frac{72}{(n-1)(n-2)}$.

Now for arrays (b) to (e), which have all the same numerical factor and the same pattern function and can therefore be considered together. For any one the numerical factor is

$$\left(\frac{4!}{2!1!1!}\right)\left(\frac{4!}{3!1!}\right)\left(\frac{2!}{1!1!}\right).\frac{1}{2!} = 48$$

and that of the four together is thus 192.

Under Rule 6 the pattern function will depend on the configuration

$$\begin{matrix} \times & \times & \times \\ \times & \times & . \\ \times & . & \times \end{matrix}$$

where $\times$ stands for a non-zero entry and a period for a zero entry. There are five separations of this, one of one separate, three of two separates, and one of three separates. The contribution from the first is

$$n\frac{1}{n}\frac{1}{n}\frac{1}{n} = \frac{1}{n^2}$$

for each column has a non-zero entry in the separate. The contribution from the three separations given respectively by isolating the first, second and third row will be found to be

$$n(n-1)\left[\frac{-1}{n^3(n-1)^3} + \frac{1}{n^3(n-1)^2} + \frac{1}{n^3(n-1)^2}\right] = \frac{2n-3}{n^2(n-1)^2}.$$

The contribution from the separation of three separates is

$$n(n-1)(n-2)\left[\frac{2!}{n(n-1)(n-2)}\ \frac{-1}{n(n-1)}\ \frac{-1}{n(n-1)}\right] = \frac{2}{n^2(n-1)^2}$$

The pattern function is the sum of these three contributions and is thus $\frac{}{1)^2}$.

The contribution from arrays (f) and (g) in (11.30) will be found to be $\dfrac{32}{(n-1)^2}$.

Hence, adding all the contributions together, we find that the coefficient of $\kappa_6\kappa_2^2$ in $\kappa(4^2 2)$ is

$$\frac{72}{(n-1)(n-2)} + \frac{192}{(n-1)^2} + \frac{32}{(n-1)^2} = \frac{8(37n-65)}{(n-1)^2(n-2)},$$

as shown in equation (11.62) below.

11.13. *Rule 10.* The expression for any $\kappa(a_1^{\pi_1} \ldots)$ which contains a unit part may be obtained from that without the part by (1) dividing throughout by n and (2) increasing the suffix of one of the κ's by unity in every possible way.

For example, it may be shown that

$$\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}.$$

Hence
$$\kappa(2^2 1) = \frac{\kappa_5}{n^2} + \frac{4\kappa_3\kappa_2}{n(n-1)}$$

$$\kappa(2^2 1^2) = \frac{\kappa_6}{n^3} + \frac{4\kappa_3^2}{n^2(n-1)} + \frac{4\kappa_4\kappa_2}{n^2(n-1)},$$

and so on.

11.14. The reader may be inclined to doubt whether this rather elaborate combinatorial procedure represents much of an advance on the straightforward algebraical approach considered earlier in the chapter. A few trials of the two methods in particular cases will soon convert him to the former. The division of the coefficients into a numerical factor and a pattern function greatly simplifies the method and in fact all the functions likely to be required for practical purposes have been tabulated by Fisher (1928) or can be derived therefrom by an iterative process given by Fisher and Wishart (cf. Exercise 11.11).

Example 11.2

To find the variance of the second moment statistic m_2.

From (11.23) we have

$$k_2 = \frac{n}{n-1} m_2.$$

Hence
$$\text{var } m_2 = \left(\frac{n-1}{n}\right)^2 \text{var } k_2$$

$$= \left(\frac{n-1}{n}\right)^2 \kappa(2^2).$$

$\kappa(2^2)$ consists of two terms, one in κ_4 and one in κ_2^2. The only array contributing to the first is

$$\begin{array}{cc|c} 2 & 2 & 4 \\ \hline 2 & 2 & 4 \end{array}$$

for the second, and we divide by 2! since there are 2 members of the row totals the same, this being the only β factor.

Under Rule 8, the pattern function is $\frac{1}{n}$ times that of

$$\begin{array}{cc} \times & \times \\ \times & \times \\ \times & \times \end{array}$$

There are five separations of this, one of one separate, three of two separates and one of three separates. The contributions respectively under Rule 6 will be found to be

$$n \cdot \left(\frac{1}{n}\right)\left(\frac{1}{n}\right) = \frac{1}{n}$$

$$3n(n-1)\left\{\frac{}{n(n-1)}\right\}\left\{\frac{-1}{n(n-1)}\right\} \qquad \frac{}{n(n-1)}$$

$$n(n-1)(n-2)\left\{\frac{(-1)^2 2!}{n(n-1)(n-2)}\right\}\left\{\frac{(-1)^2 2!}{n(n-1)(n-2)}\right\} \qquad \frac{}{n(n-1)(n-2)}.$$

The sum of these is $\frac{n}{(n-1)(n-2)}$ and hence the contribution from array (α) in (11.30) is $\frac{72}{(n-1)(n-2)}$.

Now for arrays (b) to (e), which have all the same numerical factor and the same pattern function and can therefore be considered together. For any one the numerical factor is

$$\left(\frac{4!}{2!\,1!\,1!}\right)\left(\frac{4!}{3!\,1!}\Big/\frac{2!}{1!\,1!}\right)\cdot\frac{1}{2!} = 48$$

and that of the four together is thus 192.

Under Rule 6 the pattern function will depend on the configuration

$$\begin{array}{ccc} \times & \times & \times \\ \times & \times & \cdot \\ \times & \cdot & \times \end{array}$$

where $\times$ stands for a non-zero entry and a period for a zero entry. There are five separations of this, one of one separate, three of two separates, and one of three separates. The contribution from the first is

$$n\frac{1}{n}\frac{1}{n}\frac{1}{n} = \frac{1}{n^2}$$

for each column has a non-zero entry in the separate. The contribution from the three separations given respectively by isolating the first, second and third row will be found to be

$$n(n-1)\left[\frac{-1}{n^3(n-1)^3} + \frac{1}{n^3(n-1)^2} + \frac{1}{n^3(n-1)^2}\right] = \frac{2n-3}{n^2(n-1)^2}.$$

The contribution from the separation of three separates is

$$n(n-1)(n-2)\left[\frac{2!}{n(n-1)(n-2)}\,\frac{-1}{n(n-1)}\,\frac{-1}{n(n-1)}\right] = \frac{2}{n^2(n-1)^2}.$$

The pattern function is the sum of these three contributions and is thus $\frac{1}{(n-1)^2}$.

The contribution from arrays (f) and (g) in (11.30) will be found to be $\dfrac{32}{(n-1)^2}$.

Hence, adding all the contributions together, we find that the coefficient of $\kappa_8\kappa_2^2$ in $\kappa(4^22)$ is

$$\frac{72}{(n-1)(n-2)} + \frac{192}{(n-1)^2} + \frac{32}{(n-1)^2} = \frac{8(37n-65)}{(n-1)^2(n-2)},$$

as shown in equation (11.62) below.

11.13. *Rule 10.* The expression for any $\kappa(a_1^{\alpha_1} \ldots)$ which contains a unit part may be obtained from that without the part by (1) dividing throughout by n and (2) increasing the suffix of one of the κ's by unity in every possible way.

For example, it may be shown that

$$\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}.$$

Hence

$$\kappa(2^21) = \frac{\kappa_5}{n^2} + \frac{4\kappa_3\kappa_2}{n(n-1)}$$

$$\kappa(2^21^2) = \frac{\kappa_6}{n^3} + \frac{4\kappa_3^2}{n^2(n-1)} + \frac{4\kappa_4\kappa_2}{n^2(n-1)},$$

and so on.

11.14. The reader may be inclined to doubt whether this rather elaborate combinatorial procedure represents much of an advance on the straightforward algebraical approach considered earlier in the chapter. A few trials of the two methods in particular cases will soon convert him to the former. The division of the coefficients into a numerical factor and a pattern function greatly simplifies the method and in fact all the functions likely to be required for practical purposes have been tabulated by Fisher (1928) or can be derived therefrom by an iterative process given by Fisher and Wishart (cf. Exercise 11.11).

Example 11.2

To find the variance of the second moment statistic m_2.

From (11.23) we have

$$k_2 = \frac{n}{n-1}m_2.$$

Hence

$$\operatorname{var} m_2 = \left(\frac{n-1}{n}\right)^2 \operatorname{var} k_2$$

$$= \left(\frac{n-1}{n}\right)^2 \kappa(2^2).$$

$\kappa(2^2)$ consists of two terms, one in κ_4 and one in κ_2^2. The only array contributing to the first is

$$\begin{array}{cc|c} 2 & 2 & 4 \\ \hline 2 & 2 & 4 \end{array}$$

with a numerical factor unity and a pattern function $\frac{1}{n}$. The arrays giving the second are of type

$$2 \quad 2$$

If any entry in this were a 2 the row in which it appeared would contain only a single entry and hence the array would vanish. The only contributing array is therefore

$$\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 1 & 2 \\ \\ 2 & 2 & 4 \end{array}$$

The numerical coefficient is $\left(\frac{2!}{1!1!}\right)^2 \cdot \frac{1}{2!} = 2$. The pattern function will be found to be $\frac{1}{(n-1)}$. Hence

$$\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}$$

$$\begin{aligned} \text{var } m_2 &= \left(\frac{n-1}{n}\right)^2 \kappa(2^2) \\ &= \left(\frac{n-1}{n}\right)^2 \left\{\frac{1}{n}(\mu_4 + 3\mu_2^2) + \frac{2}{n-1}\mu_2^2\right. \\ &\quad \frac{(n-1)^2}{n^3}\mu_4 + \frac{(3-n)(n-1)}{n^3}\mu_2^2. \end{aligned}$$

As n becomes large this result tends to

$$\frac{1}{n}(\mu_4 - \mu_2^2),$$

confirming the approximation given by equation (9.9).

Example 11.3

To find the third moment of k_2 we require $\kappa(2^3)$. This will be the sum of factors in κ_6, $\kappa_4\kappa_2$, κ_3^2 and κ_2^3.

The coefficient of the first is $\frac{1}{n}$. For the second we have to consider the array

$$\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ 1 & 1 & . & 2 \\ \hline 2 & 2 & 2 & \end{array}$$

all others vanishing except the two equivalent partitions obtained when the column with the single entry appears in the first or second place. The numerical factor is then

$$3.\left(\frac{2!}{1!\,1!}\right)^2 = 12.$$

The pattern function is $\frac{1}{n}$ times that of

$$\begin{matrix} \times & \times \\ \times & \times \end{matrix}$$

i.e. is $\frac{1}{n(n-1)}$. The coefficient of $\kappa_4\kappa_2$ is then $\frac{12}{n(n-1)}$.

For the term in κ_3^2 the only contributory array is

$$\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 1 & 1 & 3 \\ \\ 2 & 2 & 2 & 6 \end{array}$$

with a factor $\left(\frac{2!}{1!\,1!}\right)^3 \cdot \frac{1}{2!} = 4$ and pattern function $\frac{n-2}{n(n-1)^2}$.

For the last term we have to consider the array

$$\begin{array}{ccc|c} 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ . & 1 & 1 & 2 \\ \\ 2 & 2 & 2 & 6 \end{array}$$

with a numerical coefficient 8 and a pattern function $\frac{1}{(n-1)^2}$. Collecting terms together we get

$$\kappa(2^3) = \frac{\kappa^6}{n} + \frac{12\kappa_4\kappa_2}{n(n-1)} + \frac{4(n-2)}{n(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2}\kappa_2^3.$$

This is also the value of the third moment $\mu(2^3)$ measured about the mean of the sampling distribution κ_2. We see that if the parent is normal the third moment reduces to $\frac{8\kappa_2^3}{(n-1)^2}$, i.e. is of order n^{-2}, indicating a rapid tendency towards symmetry.

Example 11.4

Few things illustrate the usefulness of expressing the formulae in terms of cumulants and the power of the combinatorial method better than the simplification imported when the parent population is normal. In this case only terms in κ_2 survive, all higher cumulants vanishing.

As an illustration let us prove that $\kappa(pq) = 0$ for normal samples unless $p = q$.

The only term which can appear in $\kappa(pq)$ is $\kappa_2^{\frac{1}{2}(p+q)}$ and evidently, if $p - q$ is odd, even this cannot do so. If $p + q$ is even we have to consider the array

$$\begin{array}{cc|c} p & q & p+q \end{array}$$

Now if any entry in this array is 2 the array vanishes since the row concerned will contain only one entry. The reverse can only happen if all the entries are unity, in which case the sums p and q must be equal. This establishes the result.

Example 11.5

Any $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ containing π parts is of order $n^{-(\pi-1)}$. For example, $\kappa(3^2 2^2)$ is of order n^{-3}.

To prove this result we have to consider only the pattern function. Consider the array

$$\begin{array}{c|c} & a \\ \hline a_1 \quad a_1 \ldots a_s & a \end{array}$$

To the single separate there corresponds under Rule 6 the function $n\left(\dfrac{1}{n}\right)^{\pi} = n^{-(\pi-1)}$. Furthermore, no pattern function can be of greater order in n; for in an array with more than one row, with q separates there is associated the factor $n^{[q]}\dfrac{1}{n^{[\rho_1]}}\cdot\dfrac{1}{n^{[\rho_2]}} \ldots \dfrac{1}{n^{[\rho_\pi]}}$, where ρ_1 is the number of separates in the first column containing a non-zero entry, and so on. If there is more than one entry in the jth column the factor $n^{[\rho_j]}$ must be at least of order n^2 and thus the pattern function of order less than $n^{-(\pi-1)}$; and if there is only one entry in each column the order must be $n^{-(\pi-1)}$ unless the function vanishes. Hence the result.

11.15. By the above methods Professor Fisher worked out the sampling formulae for degree not greater than 10, and gave some of the 12th degree. The following are the results, with a number of corrections.

Second k-Statistic

$$\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1} \tag{11.31}$$

$$\kappa(2^3) = \frac{\kappa_6}{n^2} + \frac{12\kappa_4\kappa_2}{n(n-1)} + \frac{4(n-2)}{n(n-1)^2}\,[\text{illegible}] + \frac{[\text{illegible}]}{(n-1)^2}\,[\text{illegible}] \tag{11.32}$$

$$\begin{aligned}(2^4) = {} & \frac{\kappa_8}{n^3} + \frac{24}{n^2(n-1)}\kappa_6\kappa_2 + \frac{32(n-2)}{n^2(n-1)^2}\kappa_5\kappa_3 + \frac{8(4n^2-9n+6)}{n^2(n-1)}\,[\text{illegible}] \\ & + \frac{144}{n(n-1)^2}\kappa_4\kappa_2^2 + \frac{96(n-2)}{n(n-1)^3}\kappa_3^2\kappa_2 + \frac{48}{(n-1)^3}\kappa_2^4\end{aligned} \tag{11.33}$$

$$\begin{aligned}\kappa(2^5) = {} & \frac{\kappa_{10}}{n^4} + \frac{40\kappa_8\kappa_2}{n^3(n-1)} + \frac{80(n-2)}{n^3(n-1)^2}\kappa_7\kappa_3 + \frac{40(5n^2-12n+9)}{n^3(n-1)^3}\kappa_6\kappa_4 \\ & + \frac{16(n-2)(6n^2-12n+7)}{n^3(n-1)^4}\kappa_5^2 + \frac{480}{n^2(n-1)^2}\kappa_6\kappa_2^2 + \frac{1280(n-2)}{n^2(n-1)^3}\kappa_5\kappa_3\kappa_2 \\ & + \frac{320(4n^2-9n+6)}{n^2(n-1)^4}\kappa_4^2\kappa_2 + \frac{480(2n^2-7n+6)}{n^2(n-1)^4}\kappa_4\kappa_3^2 + \frac{1920}{n(n-1)^3}\,[\text{illegible}] \\ & + \frac{1920(n-2)}{n(n-1)^4}\kappa_3^2\kappa_2^2 + \frac{384}{(n-1)^4}\kappa_2^5\end{aligned} \tag{11.34}$$

$$
\begin{aligned}
\kappa(2^6) = {} & \frac{1}{n^5}\kappa_{12} + \frac{60}{n^4(n-1)}\kappa_{10}\kappa_2 + \frac{160(n-2)}{n^4(n-1)^2}\kappa_9\kappa_3 + \frac{240(2n^2-5n-4)}{n^4(n-1)^3}\kappa_8\kappa_4 \\
& + \frac{96(n-2)(7n^2-14n+9)}{n^4(n-1)^4}\kappa_7\kappa_5 + \frac{4(113n^4-520n^3+950n^2-800n+265)}{n^4(n-1)^5}\kappa_6^2 \\
& + \frac{1200}{n^3(n-1)^2}\kappa_8\kappa_2^2 + \frac{4800(n-2)}{n^3(n-1)^3}\kappa_7\kappa_3\kappa_2 + \frac{2400(5n^2-12n+9)}{n^3(n-1)^4}\kappa_6\kappa_4\kappa_2 \\
& + \frac{160(n-2)(31n-53)}{n^3(n-1)^4}\kappa_6\kappa_3^2 + \frac{960(n-2)(6n^2-12n+7)}{n^3(n-1)^5}\kappa_5^2\kappa_2 \\
& + \frac{1920(n-2)(9n^2-23n+16)}{n^3(n-1)^5}\kappa_5\kappa_4\kappa_3 + \frac{480(11n^3-41n^2+59n-31)}{n^3(n-1)^5}\kappa_4^3 \\
& + \frac{9600}{n^2(n-1)^3}\kappa_6\kappa_2^3 + \frac{38400(n-2)}{n^2(n-1)^4}\kappa_5\kappa_3\kappa_2^2 + \frac{9600(4n^2-9n+6)}{n^2(n-1)^4}\kappa_4^2\kappa_2^2 \\
& + \frac{28800(2n^2-7n+6)}{n^2(n-1)^5}\kappa_4\kappa_3^2\kappa_2 + \frac{960(n-2)(5n-12)}{n^2(n-1)^5}\kappa_3^4 + \frac{28800}{n(n-1)^4}\kappa_4\kappa_2^4 \\
& + \frac{38400(n-2)}{n(n-1)^5}\kappa_3^2\kappa_2^3 + \frac{3840}{(n-1)^5}\kappa_2^6 \quad . \qquad (11.35)
\end{aligned}
$$

Third k-Statistic

$$
\kappa(3^2) = \frac{1}{n}\kappa_6 + \frac{9}{n-1}\kappa_4\kappa_2 + \frac{9}{n-1}\kappa_3^2 + \frac{6n}{(n-1)(n-2)}\kappa_2^3 \quad . \qquad (11.36)
$$

$$
\begin{aligned}
\kappa(3^3) = {} & \frac{1}{n^2}\kappa_9 + \frac{27}{n(n-1)}\kappa_7\kappa_2 + \frac{27(3n-4)}{n(n-1)^2}\kappa_6\kappa_3 + \frac{27(4n-7)}{n(n-1)^2}\kappa_5\kappa_4 \\
& + \frac{54(4n-7)}{(n-1)^2(n-2)}\kappa_5\kappa_2^2 + \frac{162(5n-12)}{(n-1)^2(n-2)}\kappa_4\kappa_3\kappa_2 + \frac{36(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_3^3 \\
& + \frac{108n(5n-12)}{(n-1)^2(n-2)^2}\kappa_3\kappa_2^3 \quad . \qquad (11.37)
\end{aligned}
$$

$$
\begin{aligned}
\kappa(3^4) = {} & \frac{1}{n^3}\kappa_{12} + \frac{54}{n^2(n-1)}\kappa_{10}\kappa_2 + \frac{108(2n-3)}{n^2(n-1)^2}\kappa_9\kappa_3 + \frac{27(17n^2-49n+35)}{n^2(n-1)^3}\kappa_8\kappa_4 \\
& + \frac{108(7n^2-20n+16)}{n^2(n-1)^3}\kappa_7\kappa_5 + \frac{27(17n^2-47n+39)}{n^2(n-1)^3}\kappa_6^2 + \frac{27(37n-70)}{n(n-1)^2(n-2)}\kappa_8\kappa_2^2 \\
& + \frac{324(19n^2-67n+54)}{n(n-1)^3(n-2)}\kappa_7\kappa_3\kappa_2 + \frac{162(65n^2-245n+234)}{n(n-1)^2(n-2)}\kappa_6\kappa_4\kappa_2 \\
& + \frac{108(82n^3-481n^2+958n-640)}{n(n-1)^3(n-2)^2}\kappa_6\kappa_3^2 + \frac{108(59n^2-220n+224)}{n(n-1)^3(n-2)}\kappa_5^2\kappa_2 \\
& + \frac{324(75n^3-473n^2+1016n-756)}{n(n-1)^3(n-2)^2}\kappa_5\kappa_4\kappa_3 \\
& + \frac{27(173n^4-1503n^3+4962n^2-7380n+4200)}{n(n-1)^3(n-2)^3}\kappa_4^3 \\
& + \frac{108(71n^2-263n+234)}{(n-1)^3(n-2)^2}\kappa_6\kappa_2^3 + \frac{648(79n^2-343n+378)}{(n-1)^3(n-2)^2}\kappa_5\kappa_3\kappa_2^2 \\
& + \frac{486(63n^2-290n+352)}{(n-1)^3(n-2)^2}\kappa_4^2\kappa_2^2 + \frac{972(99n^3-688n^2+1612n-1280)}{(n-1)^3(n-2)^3}\kappa_4\kappa_3^2\kappa_2
\end{aligned}
$$

$$\frac{162(87n^3 - 594n^2 + 1420n - 1176)}{(n-1)^3(n-2)^3}\kappa_3^4 + \frac{972(23n^2 - 103n + 118)}{(n-1)^3(n-2)^3}\kappa_4\kappa_2^4$$

$$+\frac{648n(103n^2 - 510n + 640)}{(n-1)^3(n-2)^3}\kappa_3^2\kappa_2^3 + \frac{648n^2(5n-12)}{(n-1)^3(n-2)^3}\kappa_2^6 \quad . \quad . \quad (11.38)$$

Fourth k-Statistic

$$\kappa(4^2) = \frac{1}{n}\kappa_8 + \frac{16}{n-1}\kappa_6\kappa_2 + \frac{48}{n-1}\kappa_5\kappa_3 + \frac{34}{n-1}\kappa_4^2 + \frac{72n}{(n-1)(n-2)}\kappa_4\kappa_2^2$$

$$+ \frac{144n}{(n-1)(n-2)}\kappa_3^2\kappa_2 + \frac{24n(n+1)}{(n-1)(n-2)(n-3)}\kappa_2^4 \quad . \quad . \quad (11.39)$$

$$\kappa(4^3) = \frac{1}{n^2}\kappa_{12} + \frac{48}{n(n-1)}\kappa_{10}\kappa_2 + \frac{16(13n-17)}{n(n-1)^2}\kappa_9\kappa_3 + \frac{12(41n-65)}{n(n-1)^2}\kappa_8\kappa_4$$

$$+ \frac{48(16n-29)}{n(n-1)^2}\kappa_7\kappa_5 + \frac{12(37n-70)}{n(n-1)^2}\kappa_6^2 + \frac{72(11n-19)}{(n-1)^2(n-2)}\kappa_8\kappa_2^2$$

$$+ \frac{288(19n-41)}{(n-1)^2(n-2)}\kappa_7\kappa_3\kappa_2 + \frac{48(203n-523)}{(n-1)^2(n-2)}\kappa_6\kappa_4\kappa_2$$

$$+\frac{144(56n^2 - 257n + 302)}{(n-1)^2(n-2)^2}\kappa_6\kappa_3^2 + \frac{1440(4n-11)}{(n-1)^2(n-2)}\kappa_5^2\kappa_2$$

$$+ \frac{1152(22n^2 - 106n + 133)}{(n-1)^2(n-2)^2}\kappa_5\kappa_4\kappa_3 + \frac{8(709n^2 - 3430n + 4456)}{(n-1)^2(n-2)^2}\kappa_4^3$$

$$+ \frac{288(19n^3 - 98n^2 + 125n + 2)}{(n-1)^2(n-2)^2(n-3)}\kappa_6\kappa_2^3 + \frac{1728(24n^3 - 140n^2 + 200n + 4)}{(n-1)^2(n-2)^2(n-3)}\kappa_5\kappa_3\kappa_2^2$$

$$+ \frac{432(49n^3 - 287n^2 + 408n + 12)}{(n-1)^2(n-2)^2(n-3)}\kappa_4^2\kappa_2^2 + \frac{864(103n^3 - 629n^2 + 948n + 24)}{(n-1)^2(n-2)^2(n-3)}\kappa_4\kappa_3^2\kappa_2$$

$$+ \frac{288(41n^4 - 384n^3 + 1209n^2 - 1282n - 36)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_3^4 + \frac{288n(53n^2 - 179n - 52)}{(n-1)^2(n-2)^2(n-3)}\kappa_4\kappa_2^4$$

$$+ \frac{1728n(29n^3 - 196n^2 + 317n + 62)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_3^2\kappa_2^3 + \frac{1728n(n+1)(n^2 - 5n + 2)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_2^6 \quad . \quad (11.40)$$

Fifth k-Statistic

$$\kappa(5^2) = \frac{1}{n}\kappa_{10} + \frac{25}{n-1}\kappa_8\kappa_2 + \frac{100}{n-1}\kappa_7\kappa_3 + \frac{200}{n-1}\kappa_6\kappa_4 + \frac{125}{n-1}\kappa_5^2$$

$$+ \frac{200n}{(n-1)(n-2)}\kappa_6\kappa_2^2 + \frac{1200n}{(n-1)(n-2)}\kappa_5\kappa_3\kappa_2 + \frac{850n}{(n-1)(n-2)}\kappa_4^2\kappa_2$$

$$+ \frac{1500n}{(n-1)(n-2)}\kappa_4\kappa_3^2 + \frac{600n(n+1)}{(n-1)(n-2)(n-3)}\kappa_4\kappa_2^3$$

$$+ \frac{1800n(n+1)}{(n-1)(n-2)(n-3)}\kappa_3^2\kappa_2^2 + \frac{120n^2(n+5)}{(n-1)(n-2)(n-3)(n-4)}\kappa_2^5 \, . \quad . \quad (11.41)$$

Sixth k-Statistic

$$\kappa(6^2) = \frac{1}{n}\kappa_{12} + \frac{1}{n-1}(36\,\kappa_{10}\kappa_2 + 180\kappa_9\kappa_3 + 465\kappa_8\kappa_4 + 780\kappa_7\kappa_5 + 461\kappa_6^2)$$
$$+ \frac{n}{(n-1)(n-2)}(450\kappa_8\kappa_2^2 + 3600\kappa_7\kappa_3\kappa_2 + 7200\kappa_6\kappa_4\kappa_2 + 6300\kappa_6\kappa_3^2$$
$$+ 4500\kappa_5^2\kappa_2 + 21600\kappa_5\kappa_4\kappa_3 + 4950\kappa_4^3)$$
$$+ \frac{n(n+1)}{(n-1)(n-2)(n-3)}(2400\kappa_6\kappa_2^3 + 21600\kappa_5\kappa_3\kappa_2^2 + 15300\kappa_4^2\kappa_2^2$$
$$+ 54000\kappa_4\kappa_3^2\kappa_2 + 8100\kappa_3^4)$$
$$+ \frac{n^2(n+5)}{(n-1)(n-2)(n-3)(n-4)}(5400\kappa_4\kappa_2^4 + 21600\kappa_3^2\kappa_2^3)$$
$$+ \frac{n(n+1)(n^2+15n-4)}{(n-1)(n-2)(n-3)(n-4)(n-5)}720\kappa_2^6 \quad . \quad . \quad . \quad . \qquad (11.42)$$

Product-Cumulant Formulae.

$$\kappa(32) = \frac{1}{n}\kappa_5 + \frac{6}{n-1}\kappa_3\kappa_2 \quad . \quad . \quad . \quad . \quad . \qquad (11.43)$$

$$\kappa(42) = \frac{1}{n}\kappa_6 + \frac{8}{n-1}\kappa_4\kappa_2 + \frac{6}{n-1}\kappa_3^2 \quad . \quad . \quad . \qquad (11.44)$$

$$\kappa(52) = \frac{1}{n}\kappa_7 + \frac{10}{n-1}\kappa_5\kappa_2 + \frac{20}{n-1}\kappa_4\kappa_3 \;. \quad . \quad . \qquad (11.45)$$

$$\kappa(62) = \frac{1}{n}\kappa_8 + \frac{12}{n-1}\kappa_6\kappa_2 + \frac{30}{n-1}\kappa_5\kappa_3 + \frac{20}{n-1}\kappa_4^2 \qquad (11.46)$$

$$\kappa(72) = \frac{1}{n}\kappa_9 + \frac{14}{n-1}\kappa_7\kappa_2 + \frac{42}{n-1}\kappa_6\kappa_3 + \frac{70}{n-1}\kappa_5\kappa_4 \qquad (11.47)$$

$$\kappa(82) = \frac{1}{n}\kappa_{10} + \frac{16}{n-1}\kappa_8\kappa_2 + \frac{56}{n-1}\kappa_7\kappa_3 \quad \frac{112}{n-1}\kappa_6\kappa_4 \quad \frac{}{n-1} \;. \qquad (11.48)$$

$$\kappa(43) = \frac{1}{n}\kappa_7 + \frac{12}{n-1}\kappa_5\kappa_2 + \frac{30}{n-1}\kappa_4\kappa_3 + \frac{36n}{(n-1)(n-2)}\kappa_3\kappa_2^2 \;. \qquad (11.49)$$

$$\kappa(53) = \frac{1}{n}\kappa_8 + \frac{1}{n-1}(15\kappa_6\kappa_2 + 45\kappa_5\kappa_3 + 30\kappa_4^2) + \frac{n}{(n-1)(n-2)}(60\kappa_4\kappa_2^2 + 90\kappa_3^2\kappa_2) \qquad (11.50)$$

$$\kappa(63) = \frac{1}{n}\kappa_9 + \frac{1}{n-1}(18\kappa_7\kappa_2 + 63\kappa_6\kappa_3 + 105\kappa_5\kappa_4)$$
$$+ \frac{n}{(n-1)(n-2)}(90\kappa_5\kappa_2^2 + 360\kappa_4\kappa_3\kappa_2 + 90\kappa_3^3) \quad . \qquad (11.51)$$

$$\kappa(73) = \frac{1}{n}\kappa_{10} + \frac{1}{n-1}(21\kappa_8\kappa_2 + 84\kappa_7\kappa_3 + 168\kappa_6\kappa_4 + 105\kappa_5^2)$$
$$+ \frac{n}{(n-1)(n-2)}(126\kappa_6\kappa_2^2 + 630\kappa_5\kappa_3\kappa_2 + 420\kappa_4^2\kappa_2 + 630\kappa_4\kappa_3^2) \quad . \quad . \qquad (11.52)$$

$$\kappa(54) = \frac{1}{n}\kappa_9 + \frac{1}{n-1}(20\kappa_7\kappa_2 + 70\kappa_6\kappa_3 + 120\kappa_5\kappa_4) + \frac{n}{(n-1)(n-2)}(120\kappa_5\kappa_2^2 + 600\kappa_4\kappa_3\kappa_2 + 180\kappa_3^3) + \frac{n(n+1)}{(n-1)(n-2)(n-3)} 240\kappa_3\kappa_2^3 \quad . \quad . \quad . \qquad (11.53)$$

$$\kappa(64) = \frac{1}{n}\kappa_{10} + \frac{1}{n-1}(24\kappa_8\kappa_2 + 96\kappa_7\kappa_3 + 194\kappa_6\kappa_4 + 120\kappa_5^2) + \frac{}{(n-1)(n-2)}(180\kappa_6\kappa_2^2 + 1080\kappa_5\kappa_3\kappa_2 + 720\kappa_4^2\kappa_2 + 1260\kappa_4\kappa_3^2) + \frac{n(n+1)}{(n-1)(n-2)(n-3)}(480\kappa_4\kappa_2^3 + 1080\kappa_3^2\kappa_2^2) \quad . \quad . \quad . \qquad (11.54)$$

$$\kappa(32^2) = \frac{1}{n^2}\kappa_7 + \frac{16}{n(n-1)}\kappa_5\kappa_2 + \frac{12(2n-3)}{n(n-1)^2}\kappa_4\kappa_3 + \frac{48}{(n-1)^2}\kappa_2^2\kappa_3 \qquad (11.55)$$

$$\kappa(42^2) = \frac{1}{n^2}\kappa_8 + \frac{20}{n(n-1)}\kappa_6\kappa_2 + \frac{8(5n-7)}{n(n-1)}\kappa_5\kappa_3 + \frac{4(7n-10)}{n(n-1)^2}\kappa_4^2 + \frac{80}{(n-1)^2}\kappa_4\kappa_2^2 + \frac{120}{(n-1)^2}\kappa_3^2\kappa_2 \qquad (11.56)$$

$$\kappa(52^2) = \frac{1}{n^2}\kappa_9 + \frac{24}{n(n-1)}\kappa_7\kappa_2 + \frac{20(3n-4)}{n(n-1)^2}\kappa_6\kappa_3 + \frac{20(5n-7)}{n(n-1)^2}\kappa_5\kappa_4 + \frac{120}{(n-1)^2}\kappa_5\kappa_2^2 + \frac{480}{(n-1)^2}\kappa_4\kappa_3\kappa_2 + \frac{120}{(n-1)^2} \qquad (11.57)$$

$$\kappa(62^2) = \frac{1}{n^2}\kappa_{10} + \frac{28}{n(n-1)}\kappa_8\kappa_2 + \frac{12(7n-9)}{n(n-1)^2}\kappa_7\kappa_3 + \frac{4(41n-56)}{n(n-1)^2}\kappa_6\kappa_4 + \frac{20(5n-7)}{n(n-1)^2}\kappa_5^2 + \frac{168}{(n-1)^2}\kappa_6\kappa_2^2 + \frac{840}{(n-1)^2}\kappa_5\kappa_3\kappa_2 + \frac{560}{(n-1)^2}\kappa_4^2\kappa_2 + \frac{840}{(n-1)^2} \qquad (11.58)$$

$$\kappa(3^22) = \frac{1}{n^2}\kappa_8 + \frac{21}{n(n-1)}\kappa_6\kappa_2 + \frac{6(8n-11)}{n(n-1)^2}\kappa_5\kappa_3 + \frac{9(3n-5)}{n(n-1)^2} + \frac{18(6n-11)}{(n-1)^2(n-2)}\kappa_4\kappa_2^2 + \frac{18(9n-20)}{(n-1)^2(n-2)}\kappa_3^2\kappa_2 + \frac{36n}{(n-1)^2(n-2)} \qquad (11.59)$$

$$\kappa(432) = \frac{1}{n^2}\kappa_9 + \frac{26}{n(n-1)}\kappa_7\kappa_2 + \frac{24(3n-4)}{n(n-1)^2}\kappa_6\kappa_3 + \frac{10(11n-17)}{n(n-1)^2}\kappa_5\kappa_4 + \frac{36(5n-9)}{(n-1)^2(n-2)}\kappa_5\kappa_2^2 + \frac{12(61n-128)}{(n-1)^2(n-2)}\kappa_4\kappa_3\kappa_2 + \frac{36(5n-12)}{(n-1)^2(n-2)}\kappa_3^3 + \frac{360n}{(n-1)^2(n-2)}\kappa_3\kappa_2^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (11.60)$$

$$\kappa(532) = \frac{1}{n^2}\kappa_{10} + \frac{31}{n(n-1)}\kappa_8\kappa_2 + \frac{101n-131}{n(n-1)^2}\kappa_7\kappa_3 + \frac{5(37n-55)}{n(n-1)^2}\kappa_6\kappa_4 + \frac{5(23n-35)}{n(n-1)^2}\kappa_5^2 + \frac{30(9n-16)}{(n-1)^2(n-2)}\kappa_6\kappa_2^2 + \frac{30(45n-92)}{(n-1)^2(n-2)}\kappa_5\kappa_3\kappa_2$$

$$+\frac{60(15n-31)}{(n-1)^2(n-2)}\kappa_4^2\kappa_2+\frac{30(45n-103)}{(n-1)^2(n-2)}\kappa_4\kappa_3^2+\frac{720n}{(n-1)^2(n-2)}\kappa_4\kappa_2^3$$
$$+\frac{1620n}{(n-1)^2(n-2)}\kappa_3^2\kappa_2^2 \quad . \tag{11.61}$$

$$\kappa(4^22)=\frac{1}{n^2}\kappa_{10}+\frac{32}{n(n-1)}\kappa_8\kappa_2+\frac{8(13n-37)}{n(n-1)^2}\kappa_7\kappa_3+\frac{4(49n-73)}{n(n-1)^2}\kappa_6\kappa_4$$
$$+\frac{4(29n-46)}{n(n-1)^2}\kappa_5^2+\frac{8(37n-65)}{(n-1)^2(n-2)}\kappa_6\kappa_2^2+\frac{1536}{(n-1)^2}\kappa_5\kappa_3\kappa_2$$
$$+\frac{144(7n-15)}{(n-1)^2(n-2)}\kappa_4^2\kappa_2+\frac{72(21n-50)}{(n-1)^2(n-2)}\kappa_4\kappa_3^2+\frac{96(10n^2-27n-1)}{(n-1)^2(n-2)(n-3)}\kappa_4\kappa_2^3$$
$$+\frac{144(17n^2-53n-2)}{(n-1)^2(n-2)(n-3)}\kappa_3^2\kappa_2^2+\frac{192(n)(n+1)}{(n-1)^2(n-2)(n-3)}\kappa_2^5 \quad . \tag{11.62}$$

$$\kappa(43^2)=\frac{\quad}{n^2}\qquad\frac{33}{n(n-1)}\qquad\frac{6(19n-25)}{n(n-1)^2}\qquad\frac{3(65n-107)}{n(n-1)^2}$$
$$+\frac{6(19n-34)}{n(n-1)^2}\kappa_5^2+\frac{18(19n-33)}{(n-1)^2(n-2)}\kappa_6\kappa_2^2+\frac{72(23n-52)}{(n-1)^2(n-2)}\kappa_5\kappa_3\kappa_2$$
$$+\frac{54(19n-48)}{(n-1)^2(n-2)}\kappa_4^2\kappa_2+\frac{54(33n^2-148n+172)}{(n-1)^2(n-2)^2}\kappa_4^2\kappa_3^2$$
$$+\frac{72n(17n-40)}{(n-1)^2(n-2)^2}\kappa_4\kappa_2^3+\frac{108n(27n-70)}{(n-1)^2(n-2)^2}\kappa_3^2\kappa_2^3+\frac{216n^2}{(n-1)^2(n-2)^2}\kappa_2^5 \quad . \tag{11.63}$$

$$\kappa(32^3)\quad\frac{1}{n^3}\kappa_9+\frac{30}{n^2(n-1)}\kappa_7\kappa_2+\frac{2(31n-53)}{n^2(n-1)^2}\kappa_6\kappa_3+\frac{12(9n^2-23n+16)}{n^2(n-1)^3}\kappa_5\kappa_4$$
$$+\frac{240}{n(n-1)^2}\kappa_5\kappa_2^2+\frac{360(2n-3)}{n(n-1)^3}\kappa_4\kappa_3\kappa_2+\frac{24(5n-12)}{n(n-1)^3}\kappa_3^3+\frac{480}{(n-1)^3}\kappa_3\kappa_2^3 \tag{11.64}$$

$$\kappa(42^3)=\frac{1}{n^3}\kappa_{10}\quad\frac{36}{n^2(n-1)}\kappa_8\kappa_2+\frac{4(23n-37)}{n^2(n-1)^2}\qquad\frac{4(47n^2-120n+81)}{n^2(n-1)^3}\kappa_6\kappa_4$$
$$+\frac{12(9n^2-24n+17)}{n^2(n-1)^3}\kappa_5^2+\frac{360}{n(n-1)^2}\kappa_6\kappa_2^2\qquad\frac{288(5n-7)}{n(n-1)^3}\kappa_5\kappa_3\kappa_2$$
$$\frac{144(7n-10)}{n(n-1)^3}\qquad+\frac{24(49n-95)}{n(n-1)^3}\kappa_4\kappa_3^2+\frac{960}{(n-1)^3}\kappa_4\kappa_2^3+\frac{2160}{(n-1)^3}\kappa_3^2\kappa_2^2 \tag{11.65}$$

$$\kappa(3^22^2)=\ 1\qquad\frac{37}{n^2(n-1)}\qquad\frac{6(17n-27)}{n^2(n-1)^2}\qquad\frac{3(61n^2-166n+117)}{n^2(n-1)^3}\kappa_6\kappa_4$$
$$+\frac{2(59n^2-154n+113)}{n^2(n-1)^3}\kappa_5^2+\frac{6(67n-131)}{n(n-1)^2(n-2)}\kappa_6\kappa_2^2$$
$$+\frac{24(71n^2-246n+202)}{n(n-1)^3(n-2)}\qquad\frac{36(29n^2-103n+93)}{n(n-1)^3(n-2)}$$
$$+\frac{36(38n^2-155n+160)}{n(n-1)^3(n-2)}\kappa_4\kappa_3^2\qquad\frac{72(14n-23)}{(n-1)^3(n-2)}\kappa_4\kappa_2^3$$
$$\frac{144(19n-44)}{(n-1)^3(n-2)}\qquad\frac{288n}{(n-1)^3(n-2)}\kappa_2^5 \tag{11.66}$$

11.16. Additional formulae for the case of a normal parent population have been worked out by Wishart (1930). There are two general formulae :—

$$\kappa(2^r) = \frac{2^{r-1}(r-1)!}{(n-1)^{r-1}}\kappa_2{}^r \quad . \quad . \quad . \qquad (11.67)$$

$$\kappa(p^q 2^r) = \frac{2^r(r+\frac{1}{2}pq-1)!}{(\frac{1}{2}pq-1)!\,(n-1)^r}\kappa_2{}^r\kappa(p^q) \quad . \qquad (11.68)$$

and the following specific formulae of degree 12 and upwards (those of degree 10 and lower, of course, being derivable from equations (11.30) to (11.66) by putting all κ's higher than the second equal to zero).

$$\kappa(3^2 2^4) = \frac{34{,}560n}{(n-1)^5(n-2)}\kappa_2^7 \quad . \quad . \qquad (11.69)$$

$$\kappa(3^4 2) = \frac{7776n^2(5n-12)}{(n-1)^4(n-2)^3}\kappa_2^7 \quad . \quad . \qquad (11.70)$$

$$\kappa(3^4 2^2) = \frac{108{,}864n^2(5n-12)}{(n-1)^5(n-2)^3}\kappa_2^8 \quad . \qquad (11.71)$$

$$\kappa(3^4 2^3) = \frac{1{,}741{,}824n^2(5n-12)}{(n-1)^6(n-2)^3}\kappa_2^9 \quad . \qquad (11.72)$$

$$\kappa(3^6) = \frac{466{,}560n^3(22n^2-111n+142)}{(n-1)^5(n-2)^5}\kappa_2^9 \qquad (11.73)$$

$$\kappa(3^6 2) = \frac{18}{n-1}\kappa_2\kappa(3^6) \quad . \quad . \qquad (11.74)$$

$$\kappa(3^6 2^2) = \frac{360}{(n-1)^2}\kappa_2^2\kappa(3^6) \quad . \quad . \qquad (11.75)$$

$$\kappa(4^2 2^2) = \frac{1920n(n+1)}{(n-1)^3(n-2)(n-3)}\kappa_2^6 \qquad (11.76)$$

$$\kappa(4^2 2^3) = \frac{23{,}040n(n+1)}{(n-1)^4(n-2)(n-3)}\kappa_2^7 \qquad (11.77)$$

$$\kappa(4^2 2^4) = \frac{322{,}560n(n+1)}{(n-1)^5(n-2)(n-3)}\kappa_2^8 \quad . \qquad (11.78)$$

$$\kappa(4^3 2) = \frac{20{,}736n(n+1)(n^2-5n+2)}{(n-1)^3(n-2)^2(n-3)^2}\kappa_2^7 \qquad (11.79)$$

$$\kappa(4^3 2^2) = \frac{290{,}304n(n+1)(n^2-5n+2)}{(n-1)^4(n-2)^2(n-3)^2}\kappa_2^8 \qquad (11.80)$$

$$\kappa(4^3 2^3) = \frac{4{,}644{,}864n(n+1)(n^2-5n+2)}{(n-1)^5(n-2)^2(n-3)^2}\kappa \qquad (11.81)$$

$$\kappa(4^4) = \frac{6912n(n+1)}{(n-1)^3(n-2)^3(n-3)^3}\{53n^4-428n^3+1025n^2-474n+180\}\kappa_2^8. \quad (11.82)$$

$$\kappa(4^4 2) = \frac{16}{n-1}\kappa_2\kappa(4^4) \quad . \quad . \quad . \quad . \qquad (11.83)$$

$$\kappa(4^4 2^2) = \frac{288}{(n-1)^2}\kappa_2^2\kappa(4^4) \quad . \quad . \quad . \quad . \qquad (11.84)$$

$$\kappa(4^5) = \frac{484 . 12^5}{n^4}\kappa_2^{10} \text{ approximately} \quad . \quad . \qquad (11.85)$$

In virtue of the result of Example 11.4, expressions of odd degree vanish, e.g. $\kappa(32^r) = \kappa(52^r) = 0$. Further, in virtue of (11.68), $\kappa(p2^r) = 0$ if $p > 2$, for $\kappa(p) = \kappa_p = 0$ for the normal distribution. Methods of proof of (11.67) and (11.68) are suggested in Exercise 11.9. Exact results for $\kappa(4^5)$ and $\kappa(4^6)$ are given by Hsu and Lawley (1939).

Proof of the Validity of the Rules

11.17. We now proceed to prove the validity of the rules enunciated and exemplified above. Rules 1 and 2 have already been proved.

As a preliminary let us define an operator ∂_p such that

$$\left.\begin{aligned} \partial_p \mu'_r &= r(r-1)\ldots(r-p+1)\mu'_{r-p} \qquad & r > p \\ \partial_p \mu'_p &= p! & \\ \partial_p \mu'_r &= 0 & r < p \end{aligned}\right\} \qquad . \quad . \quad (11.86)$$

and
$$\partial_p(AB) = (\partial_p A)B + A(\partial_p B) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.87)$$

so that ∂ acting on a product is distributive.

In virtue of (11.87) we have

$$\begin{aligned} \partial_p(\mu'_r)^m &= m(\mu'_r)^{m-1}\,\partial_p\mu'_r \\ &= \frac{\partial}{\partial\mu'_r}(\mu'_r)^m\,\partial_p\mu'_r. \end{aligned}$$

It follows that if f is a polynomial function in the μ's

$$\partial_p f = \frac{\partial f}{\partial\mu'_1}\partial_p\mu'_1 + \frac{\partial f}{\partial\mu'_2}\partial_p\mu'_2 + \quad . \quad . \quad . \quad . \quad (11.88)$$

and this also holds if f can be expanded in a series of polynomials in the μ's.

Now consider the expression defining the seminvariants in terms of the moments (3.11):

$$\exp\left(\kappa_1 t + \ldots + \kappa_p\frac{t^p}{p!} + \ldots\right) = 1 + \mu'_1 t + \ldots \frac{\mu'_p t^p}{p!}$$

On operating on both sides by ∂_p there results

$$\begin{aligned} \exp\left(\kappa_1 t + \ldots + \kappa_p\frac{t^p}{p!} + \ldots\right)\left(\partial_p\kappa_1 t + \ldots + \partial_p\kappa_p\frac{t^p}{p!} + \ldots\right) &= t^p + \mu'_1 t^{p+1} + \ldots \\ &= t^p\left(1 + \mu'_1 t + \ldots + \mu'_p\frac{t^p}{p!} + \right. \end{aligned}$$

and hence

$$\partial_p\kappa_1 t + \ldots + \partial_p\kappa_p\frac{t^p}{p!}$$

This is an identity in t and hence

$$\left.\begin{aligned} \partial_p\kappa_p &= p! \\ \partial_q\kappa_p &= 0 \qquad q \neq p \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (11.89)$$

For example,

$$\begin{aligned} \kappa_4 &= \mu'_4 - 4\mu'_3\mu'_1 - 3\mu'^2_2 + 12\mu'_2\mu'^2_1 - 6\mu'^4_1 \\ \partial_1\kappa_4 &= 4\mu'_3 - 4\mu'_3 - 12\mu'_2\mu'_1 - 12\mu'_2\mu'_1 + 24\mu'^3_1 + 24\mu'_2\mu'_1 - 24\mu'^3_1 \\ &= 0 \\ \partial_2\kappa_4 &= 12\mu'_2 - 24\mu'^2_1 - 12\mu'_2 + 24\mu'^2_1 \\ &= 0 \\ \partial_3\kappa_4 &= 24\mu'_1 - 24\mu'_1 \\ &= 0 \\ \partial_4\kappa_4 &= 4! \end{aligned}$$

11.16. Additional formulae for the case of a normal parent population have been worked out by Wishart (1930). There are two general formulae :—

$$\kappa(2^r) = \frac{2^{r-1}(r-1)!}{(n-1)^{r-1}}\kappa_2^r \tag{11.67}$$

$$\kappa(p^q 2^r) = \frac{2^r(r+\frac{1}{2}pq-1)!}{(\frac{1}{2}pq-1)!\,(n-1)^r}\kappa_2^r\kappa(p^q) \;. \tag{11.68}$$

and the following specific formulae of degree 12 and upwards (those of degree 10 and lower, of course, being derivable from equations (11.30) to (11.66) by putting all κ's higher than the second equal to zero).

$$\kappa(3^2 2^4) = \frac{34{,}560n}{(n-1)^5(n-2)}\kappa_2^7 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{11.69}$$

$$\kappa(3^4 2) = \frac{7776n^2(5n-12)}{(n-1)^4(n-2)^3}\kappa_2^7 \quad . \quad . \tag{11.70}$$

$$\kappa(3^4 2^2) = \frac{108{,}864n^2(5n-12)}{(n-1)^5(n-2)^3}\kappa_2^8 \quad . \tag{11.71}$$

$$\kappa(3^4 2^3) = \frac{1{,}741{,}824n^2(5n-12)}{(n-1)^6(n-2)^3}\kappa_2^9 \quad . \tag{11.72}$$

$$\kappa(3^6) = \frac{466{,}560n^3(22n^2-111n+142)}{(n-1)^5(n-2)^5}\kappa_2^9 \tag{11.73}$$

$$\kappa(3^6 2) = \frac{18}{n-1}\kappa_2\kappa(3^6) \quad . \quad . \tag{11.74}$$

$$\kappa(3^6 2^2) = \frac{360}{(n-1)^2}\kappa_2^2\kappa(3^6) \quad . \tag{11.75}$$

$$\kappa(4^2 2^2) = \frac{1920n(n+1)}{(n-1)^3(n-2)(n-3)}\kappa_2^6 \tag{11.76}$$

$$\kappa(4^2 2^3) = \frac{23{,}040n(n+1)}{(n-1)^4(n-2)(n-3)}\kappa_2^7 \quad . \tag{11.77}$$

$$\kappa(4^2 2^4) = \frac{322{,}560n(n+1)}{(n-1)^5(n-2)(n-3)}\kappa_2^8 \quad . \tag{11.78}$$

$$\kappa(4^3 2) = \frac{20{,}736n(n+1)(n^2-5n+2)}{(n-1)^3(n-2)^2(n-3)^2}\kappa_2^7 \tag{11.79}$$

$$\kappa(4^3 2^2) = \frac{290{,}304n(n+1)(n^2-5n+2)}{(n-1)^4(n-2)^2(n-3)^2}\kappa_2^8 \tag{11.80}$$

$$\kappa(4^3 2^3) = \frac{4{,}644{,}864n(n+1)(n^2-5n+2)}{(n-1)^5(n-2)^2(n-3)^2} \;. \tag{11.81}$$

$$\kappa(4^4) = \frac{6912n(n+1)}{(n-1)^3(n-2)^3(n-3)^3}\{53n^4-428n^3+1025n^2-474n+180\}\kappa_2^8. \tag{11.82}$$

$$\kappa(4^4 2) = \frac{16}{n-1}\kappa_2\kappa(4^4) \quad . \quad . \quad . \quad . \tag{11.83}$$

$$\kappa(4^4 2^2) = \frac{288}{(n-1)^2}\kappa_2^2\kappa(4^4) \quad . \quad . \quad . \quad . \tag{11.84}$$

$$\kappa(4^5) = \frac{484 . 12^5}{n^4}\kappa_2^{10} \text{ approximately} \quad . \quad . \tag{11.85}$$

In virtue of the result of Example 11.4, expressions of odd degree vanish, e.g. $\kappa(32^r) = \kappa(52^r) = 0$. Further, in virtue of (11.68), $\kappa(p2^r) = 0$ if $p > 2$, for $\kappa(p) = \kappa_p = 0$ for the normal distribution. Methods of proof of (11.67) and (11.68) are suggested in Exercise 11.9. Exact results for $\kappa(4^5)$ and $\kappa(4^6)$ are given by Hsu and Lawley (1939).

Proof of the Validity of the Rules

11.17. We now proceed to prove the validity of the rules enunciated and exemplified above. Rules 1 and 2 have already been proved.

As a preliminary let us define an operator ∂_p such that

$$\left.\begin{aligned} \partial_p \mu'_r &= r(r-1)\ldots(r-p+1)\mu'_{r-p} \qquad r > p \\ \partial_p \mu'_p &= p! \\ \partial_p \mu'_r &= 0 \qquad r < p \end{aligned}\right\} \qquad (11.86)$$

and

$$\partial_p(AB) = (\partial_p A)B + A(\partial_p B) \quad . \quad . \quad . \qquad (11.87)$$

so that ∂ acting on a product is distributive.

In virtue of (11.87) we have

$$\begin{aligned} \partial_p(\mu'_r)^m &= m(\mu'_r)^{m-1}\,\partial_p\mu'_r \\ &= \frac{\partial}{\partial\mu'_r}(\mu'_r)^m\,\partial_p\mu'_r. \end{aligned}$$

It follows that if f is a polynomial function in the μ's

$$\partial_p f = \frac{\partial f}{\partial\mu'_1}\partial_p\mu'_1 + \frac{\partial f}{\partial\mu'_2}\partial_p\mu'_2 + \qquad (11.88)$$

and this also holds if f can be expanded in a series of polynomials in the μ's.

Now consider the expression defining the seminvariants in terms of the moments (3.11):

$$\exp\left(\kappa_1 t + \ldots + \kappa_p\frac{t^p}{p!} + \qquad = 1 + \mu'_1 t + \qquad \frac{\mu'_p t^p}{p!}\right.$$

On operating on both sides by ∂_p there results

$$\begin{aligned} \exp\left(\kappa_1 t + \ldots + \kappa_p\frac{t^p}{p!} + \ldots \quad \partial_p\kappa_1 t + \ldots \quad \partial_p\kappa_p\frac{t^p}{p!}\right. &= t^p + \mu'_1 t^{p+1} + \\ &= t^p\left(1 + \mu'_1 t + \ldots \mu'_p\frac{t^p}{p!} + \ldots\right) \end{aligned}$$

and hence

$$\partial_p\kappa_1 t + \ldots \quad + \partial_p\kappa_p\frac{t^p}{p!} + \ldots = t^p.$$

This is an identity in t and hence

$$\begin{aligned} \partial_p\kappa_p &= p! \\ \partial_q\kappa_p &= 0 \qquad \quad p) \end{aligned} \qquad . \quad (11.89)$$

For example,

$$\begin{aligned} \kappa_4 &= \mu'_4 - 4\mu'_3\mu'_1 - 3\mu'^2_2 + 12\mu'_2\mu'^2_1 - 6\mu'^4_1 \\ \partial_1\kappa_4 &= 4\mu'_3 - 4\mu'_3 - 12\mu'_2\mu'_1 - 12\mu'_2\mu'_1 + 24\mu'^3_1 + 24\mu'_2\mu'_1 - 24\mu'^3_1 \\ &= 0 \\ \partial_2\kappa_4 &= 12\mu'_2 - 24\mu'^2_1 - 12\mu'_2 + 24\mu'_1 \\ &= 0 \\ \partial_3\kappa_4 &= 24\mu'_1 - 24\mu'_1 \\ &= 0 \\ \partial_4\kappa_4 &= 4! \end{aligned}$$

11.18. Now in accordance with Rule 1, which we have already established, $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ and hence $\mu(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ may be expressed in terms of parent κ's by an equation of the form

$$\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots) = \Sigma\{A(\kappa_{b_1}^{\beta_1} \kappa_{b_2}^{\beta_2} \ldots)\} \quad . \quad . \quad . \; (11.90)$$

where A is a factor which it is our object to find. Operate on both sides of (11.90) by $(\partial_{b_1}^{\beta_1} \partial_{b_2}^{\beta_2}$ Every term on the right is annihilated except that in $(\kappa_{b_1}^{\beta_1} \kappa_{b_2}^{\beta_2}$ and we have

$$A.(b_1!)^{\beta_1} (b_2!)^{\beta_2} \ldots \beta_1! \beta_2! \ldots = (\partial_{b_1}^{\beta_1} \partial_{b_2}^{\beta_2} \ldots)\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots) \quad . \; (11.91)$$

We now consider an operator θ_p, analogous to ∂_p, which, when acting on a power of x (of any suffix), reduces the exponent by p and multiplies by $r(r-1) \ldots (r-p+1)$; and we will suppose the operator to be distributive.* Regarding $\mu(a_1^{\alpha_1} a_2$ as the mean value of $(k_{a_1}^{\alpha_1} k_{a_2}^{\alpha_2} \ldots)$ we see that the result of operating by the ∂'s on the mean value is the same as that given by taking the mean value of the operation of the θ's. But this latter operation results in a constant, which is equal to its mean value; and we thus have

$$A = \frac{(\theta_{b_1}^{\beta_1} \theta_{b_2}^{\beta_2} \ldots)}{(b_1!)^{\beta_1} (b_2!)^{\beta_2} \ldots \beta_1! \beta_2!} (k_{a_1}^{\alpha_1} k_{a_2}^{\alpha_2} \ldots). \qquad (11.92)$$

Our rules are concerned with the evaluation of this operation.

11.19. Consider now a completed array of type (11.28). A little reflection will show that there is one such array for every term in (11.92) which does not vanish by operation, and that every term in (11.92) will have its corresponding completed array. The numbers in the body of the array are the powers of x occurring in the k-product; added horizontally they compose the orders of the operators; added vertically they compose the orders of the corresponding k's. A completed array is, so to speak, a chart of part of the operation; and the whole operation is the sum of all possible completed arrays.

The operation (11.92) gives us the coefficients in $\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$, but we wish to find those in the corresponding $\kappa(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$. The necessary allowance is made by Rule 3, which we now prove; that is, the coefficient of $(\kappa_{b_1}^{\beta_1} \kappa_{b_2}^{\beta_2} \ldots)$ in $\kappa(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ is given by all completed arrays, *ignoring those which are resolvable into separate blocks each confined to separate rows and columns.*

Referring to equation (11.27), expressing the relation between multivariate moments and cumulants, we see that $\kappa(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ is the sum of terms composed of products of one, two, three . . . multivariate moments. The first term is $\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ itself. Consider a two-part term such as $\mu(a_1^{\alpha'_1} a_2^{\alpha'_2} \ldots)\mu(a_1^{\alpha''_1}$.), where $\alpha'_1 + \alpha''_1 = \alpha_1$, etc. Its coefficient in the expansion on the right-hand side of (11.27) is

$$-\tfrac{1}{2} \cdot \frac{2!}{1!\,1!} \frac{t_{a_1}^{\alpha_1}}{\alpha'_1!\,\alpha''_1!} \frac{t_{a_2}^{\alpha_2}}{\alpha'_2!\,\alpha''_2!} \ldots$$

and hence the coefficient with which it appears in the formula for $\kappa(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ is

$$\frac{\alpha_1!}{\alpha'_1!\,\alpha''_1!} \quad \alpha_2! \qquad (11.93)$$

Now $\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ will itself have an array of type (11.28) with column totals $(a_1^{\alpha'_1} a_2^{\alpha'_2} \ldots$ and row totals, say $(b_1^{\beta'_1} b_2^{\beta'_2} \ldots)$; and similarly for $\mu(a_1^{\alpha''_1} a_2^{\alpha''_2} \ldots)$. Provided that

* θ_p may be regarded as equivalent to $\left(\frac{\partial^p}{\partial x_1^p} + \frac{\partial^p}{\partial x_2^p} + \ldots \frac{\partial^p}{\partial x_n^p}\right)$, i.e. to S_p in the notation of **11.10.**

$\beta_2' + \beta_1'' = \beta_1$ these arrays will correspond to terms in the κ's which, when multiplied, will give a term in $(\kappa_{b_1}^{\beta_1}\kappa_{b_2}^{\beta_2} \ldots)$. Thus the product of these terms may be considered as an array of type (11.28) with column totals $(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ and row totals $(b_1^{\beta_1} b_2^{\beta_2} \ldots)$ and with the body of the table resolvable into two separate blocks. Since there are α_1 columns of total a_1, there will be $\binom{\alpha_1}{\alpha_1'}\binom{\alpha_2}{\alpha_2'}$ products of this type in the expression which gives $\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$. This factor is the same as (11.93) but of opposite sign. Hence, if we ignore the separate two-part blocks in the array for μ we shall have allowed for the products of two moments which must be subtracted from μ to give κ.

Now some of these separate blocks will themselves be separable into two blocks, and in subtracting them all from $\mu(a_1^{\alpha_1} a_2^{\alpha_2} \ldots)$ we subtract too much. For example, if there are three separate blocks, L, M, N, we shall, by considering L and $(M + N)$ as two blocks, have subtracted L, M, N. We shall have done the same by considering M and $(L + N)$, and N and $(L + M)$ as two blocks. That is, we have subtracted $2L, 2M, 2N$ too much. We must restore these blocks to the array for μ again. Such additions, summed over all blocks of three, will be found to equal the terms in the expansion of (11.27) which result from the product of three moments.

In restoring these blocks we restore too many of the cases where there are four separate blocks. These must be subtracted again, and correspond to the negative term in (11.27) involving the product of four moments. Proceeding in this way we establish Rule 3.

11.20. Now we proceed to Rules 4, 5 and 6, which are the fundamental rules of the whole process. Consider again the array of type (11.28) to fix the ideas, say,

$$\begin{array}{cccc} 2 & 3 & 1 & 6 \\ 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad . \quad (11.94)$$

This array will represent a number of terms in the operation each of which consists of the operation of θ_6 on a term $x^2.x^3.x$ (the first row), θ_2 on $x.x$ (the second row), and so on. Provided that the suffixes of the x's in any row are alike, every suffix of the x's will provide a term, for k_p contains terms with every distribution of powers (adding to p) and suffixes. There will, for instance, be terms of the following kind:—

$$\begin{array}{ccc} x_1^2 & x_1^3 & \\ x_2 & x_2 & \\ x_2 & . & x_3, \end{array} \qquad \begin{array}{ccc} x_1^2 & x_1^3 & \\ x_1 & x_1 & \\ x_2 & . & x_2, \end{array} \qquad x_1.$$

In fact, for any completed array, we have terms in which

(i) all the x's have the same suffix (n in number, one for each suffix).

(ii) all the x's but one row have the same suffix ($n(n-1)$ in number).

(iii) all the x's but two rows have the same suffix and the remaining two are the same ($n(n-1)$ in number),

and so on. These cases correspond to the various separations dealt with in Rule 5.

Now in case (i) the term in any column arises from the term in x^p in k_p and (apart from numerical factors which are considered presently) is n^{-1}, from equation (11.16). Hence any column which contains an entry contributes a factor n^{-1} and the total function

of n arising from case (i) is the product of n and of (n^{-1}) to the power of the number of columns containing a non-zero entry.

Similarly in cases (ii) and (iii) the n-function for each separation is the product of $n(n-1)$ and, for each column, a factor in n^{-1} or $\frac{1}{n(n-1)}$ according as the column contains non-zero entries in one or in both parts of the separation; and so on.

This explains the origin of the pattern function as described in Rule 6. But in order to establish that rule completely (and incidentally to establish Rules 4 and 5) we have to show that the numerical coefficients arising from each separation are the same. When this is done the validity of Rule 6 is demonstrated, for the separate contributions in n may be added together to give the pattern function and the whole multiplied by the numerical coefficient.

θ_1 may be considered as the operation of picking out an x from the operand in all possible ways and replacing it by unity. Similarly $\frac{\theta_p}{p!}$ may be regarded as picking out p x's with the same suffix and replacing them by unity. It is thus evident that operating on a k product by a θ product $\frac{\theta_{b_1}\theta_{b_2}}{b_1!\,b_2!\ldots}$ of the same degree will yield a result which is the number of ways in which sets of x's can be picked out of the k product so that each set contains b_1 of one suffix, b_2 of a second suffix (which may be the same as the first), and so on.

Now consider the operation (11.92) in which the k's are expressed in the simplified form (11.17). The operations θ being distributive, we shall emerge from the operation with a sum of terms comprising all the possible ways in which the individual x's can be picked out of the k product such that the row and column totals of the two-way array are satisfied. Consider the sets corresponding to a particular array, such as (11.94). The contribution to the total will consist of the ways of picking out individuals such that

(i) from the individuals in the first k_4 are chosen four in the partition (2, 1, 1),

(ii) from the second k_4 are chosen four in the partition (3, 1),

(iii) from the k_2 are chosen two in the partition (1, 1),

(iv) these are associated in all possible ways such that individuals in a row arise from the same suffix.

On consideration it will be seen that the total number of ways of doing this is the number of ways of allocating the individuals from column totals as required by Rule 5; *and this is true whether sets of rows have the same suffix or not.*

Rules 5 and 6, and hence Rule 4, follow at once.

11.21. The remaining rules are ancillary.

Rule 7 follows from Rule 2. In fact, the pattern function is independent of the numbers composing the array, and the pattern with a row containing one element can therefore form the skeleton of an array in which that element is unity; and this would entail the appearance of κ_1, which by Rule 2 is impossible.

Rule 8 follows from Rule 6. The column containing the single element appears in just one separate of all the separations, and the contributions to the pattern function are thus all multiplied by n^{-1} owing to its presence.

Rule 10 follows from Rule 8. The addition of a unit part is equivalent to the addition of an extra column containing unity. This multiplies all pattern functions by $\frac{1}{n}$, leaves

numerical coefficients unchanged and increases the suffix of every κ according to the row in which the unit appears.

11.22. There only remains to prove Rule 9. Note that any pattern function can be evaluated linearly in terms of the functions of the pattern obtained by omitting one of the columns. For example, consider the right-hand column of

$$\begin{array}{cccc} & \times & \times & \\ \times & \cdot & \times & \\ \times & & & \\ \times & \times & \cdot & \cdot \end{array} \tag{11.95}$$

and the contributions to the pattern function from it. The 15 separations which are possible with four rows can be divided into two classes, that in which the two rows in the fourth column lie in the same separate and that in which they do not. In separations of the first type the contributions from the first three columns will be the contributions of all separations of

$$\begin{array}{ccc} \times & \times & \times \\ \times & \times & \cdot \\ \times & \times & \cdot \end{array} \tag{11.96}$$

in which the first two rows are amalgamated. Considering the function of the first three rows

$$\begin{array}{ccc} & \times & \times \\ \times & \cdot & \times \\ \times & \times & \\ \times & \times & \cdot \end{array} \tag{11.97}$$

in which amalgamation has not taken place, we see that the contribution consists of all contributions which do not occur in the first. Calling the first A and the second B, we see that the contribution is

$$\frac{1}{n}A - \frac{1}{n(n-1)}(B-A) = \frac{1}{n-1}A - \frac{1}{n(n-1)}B,$$

i.e. a linear function of the derived patterns A and B. The proof of the general result follows exactly the same lines.

Now if a pattern may be divided into two groups connected only by a single column we can reduce it step by step by omitting the other columns. We end up with this single column, and the pattern function of this column must vanish; for the column total a corresponds to k_a, whose mean value the one-column array expresses, and since by definition this mean value is κ_a no composite terms such as would be given by two rows or more can appear.

11.23. As an illustration of the way in which the sampling formulae can be used to approximate to a sampling distribution, let us consider the distribution of $\sqrt{b_1}$ in samples from a normal population. We have, in terms of the sample moments,

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} = \frac{n-2}{\sqrt{n(n-1)}}\frac{k_3}{k_2^{3/2}}.$$

For a normal distribution the variance of k_3, $\kappa(3^2)$ is, by (11.36), equal to

$$\frac{6n}{(n-1)(n-2)}.$$

We therefore consider the statistic

$$x = \sqrt{\frac{(n-1)(n-2)}{6n}}\, k_3 k_2^{-\frac{3}{2}} \qquad (11.98)$$

$$= \frac{n-1}{\sqrt{\{6(n-2)\}}}\sqrt{b_1} \qquad (11.99)$$

which will, to order n^{-1}, have unit variance. We have

$$x = \sqrt{\frac{(n-1)(n-2)}{6n}}\, \frac{k_3}{\kappa_2^{\frac{3}{2}}}\left\{1 + \frac{k_2 - \kappa_2}{\kappa_2}\right\}^{-\frac{3}{2}} \qquad (11.100)$$

Since the population is symmetrical the mean value of x is zero. We then have, expanding (11.100),

$$x^2 = \frac{(n-1)(n-2)}{6n} \cdot \frac{1}{\kappa_2^3}\left\{k_3^2 - \frac{3}{\kappa_2} k_3^2(k_2 - \kappa_2) + \frac{6}{\kappa_2^2} k_3^2(k_2 - \kappa_2)^2 - \frac{10}{\kappa_2^3} k_3^2(k_2 - \kappa_2)^3\right.$$
$$\left. + \frac{15}{\kappa_2^4} k_3^2(k_2 - \kappa_2)^4 - \frac{21}{\kappa_2^5} k_3^2(k_2 - \kappa_2)^5 + \frac{28}{\kappa_2^6} k_3^2(k_2 - \kappa_2)^6 + \ldots\right\}. \qquad (11.101)$$

The variance may be obtained by taking mean values of both sides, and since κ_2 is the mean value of k_2 we have

$$\text{var}\,(x) = \frac{(n-1)(n-2)}{6n} \cdot \frac{1}{\kappa_2^3}\left\{\mu(3^2) - \frac{3}{\kappa_2^6}\mu(3^2 2) + \frac{6}{\kappa_2^2}\mu(3^2 2^2) - \frac{10}{\kappa_2^3}\mu(3^2 2^3)\right.$$
$$\left. + \frac{15}{\kappa_2^4}\mu(3^2 2^4) - \frac{21}{\kappa_2^5}\mu(3^2 2^5) + \frac{28}{\kappa_2^6}\mu(3^2 2^6) + \ldots\right\}. \qquad (11.102)$$

We now express the product μ's in terms of product κ's by using equation (11.27) and identifying coefficients. For a normal distribution $\kappa(32^r) = 0$ and we will take our approximation to order n^{-4}, so that κ's of five parts or more may be neglected. We then find,

$$\text{var}\, x = \frac{(n-1)(n-2)}{6n} \cdot \frac{1}{\kappa_2^3}\left[\kappa(3^2) - \frac{3}{\kappa_2}\kappa(3^2 2) + \frac{6}{\kappa_2^2}\{\kappa(3^2 2^2) + \kappa(3^2)\kappa(2^2)\}\right.$$
$$- \frac{10}{\kappa_2^3}\{\kappa(3^2 2^3) + 3\kappa(3^2 2)\kappa(2^2) + \kappa(3^2)\kappa(2^3)\} + \frac{15}{\kappa_2^4}\{6\kappa(3^2 2^2)\kappa(2^2) + 4\kappa(3^2 2)\kappa(2^3)$$
$$+ \kappa(3^2)\kappa(2^4) + 3\kappa(3^2)\kappa^2(2^2)\} - \frac{21}{\kappa_2^5}\{15\kappa(3^2 2)\kappa^2(2^2) + 10\kappa(3^2)\kappa(2^3)\kappa(2^2)\}$$
$$\left. + \frac{28}{\kappa_2^6} . 15\kappa(3^2)\kappa^3(2^2)\right] \qquad (11.103)$$

Substituting the values of equations (11.31) to (11.85) we find, after some purely algebraic reduction,

$$\text{var}\, x = 1 - \frac{6}{n-1} + \frac{28}{(n-1)^2} - \frac{120}{(n-1)^3}$$
$$= 1 - \frac{6}{n} + \frac{22}{n^2} - \frac{70}{n^3} + \ldots \qquad (11.104)$$

In a similar way (for details, see E. S. Pearson, 1930) we find

$$\mu_4(x) = 3 - \frac{1056}{n^2} + \frac{24{,}132}{n^3} - \ldots \qquad (11.105)$$

$\mu_3(x)$ is zero, for the distribution is symmetrical.

Thus it appears that as $n \to \infty$ the second moment of x tends to unity and the fourth moment to 3, which is in conformity with tendency to normality. But the tendency is by no means very rapid. When $n = 100$ the variance is approximately 0·942 and in assuming x to be distributed with unit variance we should commit an error of about 6 per cent.

11.24. There are two ways of improving on the first approximation that x is normally distributed with unit variance. In the first place we may consider a transformation to a new variate ξ, chosen so that ξ is normally distributed to order n^{-2}. Secondly, we may fit a Pearson curve to the distribution of x, using the values of moments given by (11.104) and (11.105). The appropriate curve is the Type VII

$$dF \propto \left(1 + \frac{x^2}{a^2}\right)^{-m} dx. \qquad \text{(11.106)}$$

The first line was adopted by Fisher (1928), who obtained the following transformation:

$$\xi = x\left(1 + \frac{3}{n} + \frac{91}{4n^2}\right) - \frac{3}{2n}\left(1 - \frac{111}{2n}\right)(x^3 - 3x) - \frac{33}{8n^2}(x^5 - 10x^3 + 15x) \qquad \text{(11.107)}$$

The second was adopted by E. S. Pearson (1930), who tabulated the 1 per cent. and 5 per cent. significance points of (11.106), that is to say the values of the deviates x for various values of n such that 99 per cent. and 95 per cent. of the total frequency of the sampling distribution falls within a range of $\pm x$ on each side of the mean.

The Multivariate Case

11.25. The foregoing results can be generalised to the multivariate case, and we give an outline of the extension to that of two variates.

Given any bipartite number pp' we shall have for any partition $\{(p_1p_1')^{\pi_1}(p_2p_2')^{\pi_2} \ldots\}$ and the bivariate cumulant $\kappa_{pp'}$ a k-statistic $k_{pp'}$ whose mean value is $\kappa_{pp'}$. Explicitly

$$k_{pp'} = p!\,p'!\,\Sigma \frac{(-1)^{\rho-1}(\rho-1)!}{n^{[\rho]}} \Sigma \frac{(x_1^{p_1}y_1^{p_1'}\,x_2^{p_1}\,y_2^{p_1'} \ldots x_\rho^{p_\rho}\,y_\rho^{p_\rho'})}{(p_1!)^{\pi_1}\,(p_1'!)^{\pi_1} \ldots \pi_1!\,\pi_2! \ldots} \qquad \text{(11.108)}$$

In particular, corresponding to (11.22) we have

$$\left.\begin{aligned}
k_{11} &= \frac{1}{n^{[2]}}(ns_{11} - s_{10}s_{01})\\
k_{21} &= \frac{1}{n^{[3]}}(n^2 s_{21} - 2ns_{10}\,s_{11} - ns_{20}\,s_{01} + 2s_{10}^2\,s_{01})\\
k_{31} &= \frac{1}{n^{[4]}}\{n^2(n+1)s_{31} - n(n+1)s_{30}\,s_{0} \quad 3n(n-1)s_{11}\,s_{2}\\
&\qquad - 3n(n+1)s_{21}\,s_{10} + 6ns_{11}\,s_{10}^2 + 6ns_{20}\,s_{10}s_{01} - 6s_{01}\,s_{10}^3\\
k_{22} &= \frac{n}{(n-1)(n-2)(n-3)}\left\{(n+1)s_{22} - \frac{2(n+1)}{n}s_{21}\,s_{01} - \frac{2(n+1)}{n}s_{12}\,s_{10}\right.\\
&\qquad \left. - \frac{2(n-1)}{n}s_{11}^2 + \frac{8}{n}s_{11}\,s_{10}\,s_{10} + \frac{2}{n}s_{02}\,s_{10}^2 + \frac{2}{n}s_{20}\,s_{01}^2 \quad \frac{6}{n^2}s_{10}^2\,s_{01}^2\right\}
\end{aligned}\right\} \qquad \text{(11.109)}$$

In generalisation of the mean value functions of the k's we may write, for example,

$$E(k_{20}\,k_{11}) = \mu\begin{pmatrix}2 & 1\\ 0 & 1\end{pmatrix}$$

$$E(k_{20}\,k_{11}\,k_{02}) = \mu\begin{pmatrix}2 & 1 & 0\\ 0 & 1 & 2\end{pmatrix}$$

with corresponding κ's. The latter may be expressed in terms of the cumulants of the bivariate distribution as in the univariate; and the coefficients will now depend on partitions of bipartite numbers. Our rules still apply (and in particular the pattern functions appropriate to particular arrays are the same); but the numerical coefficients associated with completed arrays are modified, for we now have to consider the number of ways of allocating two different sorts of individuals in a two-way partition of a bipartite number. An example will make the modification clear.

Suppose we wish to find the coefficient of $\kappa_{33}\kappa_{11}^2$ in $\kappa\begin{pmatrix}2 & 2 & 1\\ 2 & 2 & 1\end{pmatrix}$. The total degree is 10 and, the orders of the product being 6, 2, 2, we have to consider arrays of type

$$\begin{array}{ccc|c} & & & 6\\ & & & \\ & & & \\ 4 & 4 & 2 & 10\end{array}$$

i.e. those we discussed above. The pattern functions are those we have already found. For the numerical coefficients we have to regard the column totals as consisting of the two types of object in number (2, 2), (2, 2), and (1, 1) and the row totals (3, 3), (1, 1) and 1, 1). For instance, the array

$$\begin{array}{ccc|c} 2 & 2 & 2 & 6\\ 1 & 1 & . & 2\\ 1 & 1 & . & 2\\ \hline 4 & 4 & 2 & 10\end{array}$$

might be written either as

$$\begin{array}{ccc|c} (1,1) & (1,1) & (1,1) & (3,3)\\ (0,1) & (1,0) & . & (1,1)\\ (1,0) & (0,1) & . & (1,1)\\ \hline (2,2) & (2,2) & (1,1) & (5,5)\end{array} \qquad . \quad (11.110)$$

or as

$$\begin{array}{ccc|c} (2,0) & (0,2) & (1,1) & (3,3)\\ (0,1) & (1,0) & . & (1,1)\\ (0,1) & (1,0) & . & (1,1)\\ \hline (2,2) & (2,2) & (1,1) & (5,5)\end{array} \qquad . \quad (11.111)$$

each of which will make a contribution to the numerical coefficient. It will be found that no other arrays are possible except those obtained by permuting the first two columns.

The numerical coefficient in (11.110) and the permuted array together is

$$\frac{2!}{(1!\,1!)(1!\,1!)}\,\frac{2!}{(1!\,1!)(1!\,1!)}\cdot\frac{1}{2!} = 16.$$

That in (11.111) and the permuted array is

$$2\left(\frac{2!}{2!}\right)\left(\frac{2!}{1!\,1!}\right)\left(\frac{2!}{1!\,1!}\right)\left(\frac{2!}{2!}\right)\cdot\frac{1}{2!} = 4.$$

The total contribution is thus 20. The pattern function is $\dfrac{1}{(n-1)(n}$

In the same way it will be found that for the partitions

$$\begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ \\ 4 & 4 & 2 & 10 \end{array} \qquad \begin{array}{ccc|c} 3 & 3 & . & 6 \\ 1 & . & 1 & 2 \\ . & 1 & 1 & 2 \\ \\ 4 & 4 & 2 & 10 \end{array}$$

the coefficients are 48 and 8. Thus the desired coefficient of $\kappa_{33}\kappa_{11}^2$ is

$$\frac{20}{(n-1)(n-2)} + \frac{48}{(n-1)^2} + \frac{8}{(n-1)^2} = \frac{4(19n-33)}{(n-1)^2(n-2)}.$$

Example 11.6

To find an exact expression for the covariance of the estimates of variance of two correlated variables, i.e.

$$\kappa\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}.$$

This will clearly consist of three terms, in κ_{22}, $\kappa_{02}\,\kappa_{20}$ and κ_{11}^2. For the first we have the partition

$$\begin{array}{cc|c} (2,0) & (0,2) & (2,2) \\ \\ (2,0) & (0,2) & (2,2) \end{array}$$

with pattern function $\dfrac{1}{n}$ and numerical coefficient unity. For the second no contribution exists, the only arrangement being

$$\begin{array}{cc|c} (2,0) & . & (2,0) \\ . & (0,2) & (0,2) \\ \\ (2,0) & (0,2) & (2,2) \end{array}$$

which has a vanishing pattern function. For the third term we have

$$\begin{array}{cc|c} (1,0) & (0,1) & (1,1) \\ (1,0) & (0,1) & (1,1) \\ \\ (2,0) & (0,2) & (2,2) \end{array}$$

the pattern function for which is $\dfrac{1}{(n-1)}$ and numerical coefficient 2. Hence

$$\kappa\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{2}{n-1}\kappa_{11}^2$$

11.26. In conclusion it may be noted that the method of expectations may be used to derive sampling moments of the distribution of samples from a finite population. The algebra becomes much more complex because the sample values are no longer independent

and we cannot, for example, write $E(\sum_{j\neq k} x_j^r x_k^s) = \mu'_r \mu'_s$. Tschuprow (1923) and Isserlis (1931) have investigated the subject systematically, the latter giving formulae for the first four moments of the mean and the first four of the second moment. Some of these had been obtained earlier by Tschuprow himself, Neyman (1925) and Church (1926). We quote the following formulae, in which N represents the number in the population :—

Moments of the Mean :

$$E(m'_1) = \mu'_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.112)$$

$$E(m'_1 - \mu'_1)^2 = \frac{N-n}{n(N-1)}\mu_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.113)$$

$$E(m'_1 - \mu'_1)^3 = \frac{(N-n)(N-2n)}{n^2(N-1)(N-2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.114)$$

$$E(m'_1 - \mu'_1)^4 = \frac{N-n}{n^3(N-1)(N-2)(N-3)}\{(N^2 - 6Nn + N + 6n^2)\mu_4 + 3N(N-n-1)(n-1)\mu_2^2\} \quad . \quad (11.115)$$

Moments of the Sample Variance :

$$E(m_2) = \frac{(n-1)N}{n(N-1)}\mu_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.116)$$

$$E\left(m_2 - \frac{(n-1)N}{n(N-1)}\mu_2\right)^2 = \frac{N(N-n)(n-1)}{n^2(N-1)^2(N-2)(N-3)}\{(Nn - n - N - 1)(N-1)\mu_4 - (N^2 n - 3N^2 + 6N - 3n - 3)\mu_2^2\} \quad . \quad (11.117)$$

Church also gives the third and fourth moments of the sample variance, the formulae taking several pages to write down. His version of the fourth moment contains errors which were corrected by Isserlis. The statistical usefulness of these results seems to be somewhat limited, but it would be interesting to inquire how far the combinatorial method appropriate to k-statistics can be extended to the case of the finite population.

NOTES AND REFERENCES

For earlier work on the expectations of moments see Tschuprow (1919) and Church (1925). Thiele (1903) seems to have been the first to appreciate the possibilities of using other symmetric functions, but owing to the fact that he defined his sample "semi-invariants" to be the same function of the observations as the parent seminvariants (our cumulants) are of the parent values, his formulae remained complicated. Later investigations on similar lines were carried out by Craig (1929) and St. Georgescu (1932).

k-statistics were introduced by Fisher in 1928 and subsequently applied to several problems by himself (1930) and Wishart (1929*a*). The theory of the sampling of the statistics has been developed by Fisher and Wishart (1930) and applications to the normal distribution given by Wishart (1929*b* and 1930), E. S. Pearson (1930), and Hsu and Lawley (1939). The general theory of the k-statistics and other seminvariant statistics has been discussed by Kendall (1940*a*, 1940*b*, 1940*c*, 1942) and Dressel (1940). Dressel has called attention to the relationship between seminvariant statistics and the seminvariants of the theory of binary quantics. The reader who refers to Fisher's basic paper of 1928 should beware of misprints. Methods of deriving bivariate formulae from univariate formulae by symbolic processes are given by Kendall (1940*c*).

Reference is made in the text to the work by Tschuprow (1923) and Isserlis (1931) on finite populations. The latter's method appears to be the simplest known at the present time.

Church, A. E. R. (1925), "On the moments of the distributions of squared standard deviations for samples of n drawn from an indefinitely large population," *Biometrika*, **17**, 79.

—— (1926), "On the means and squared standard deviations of small samples from any population," *Biometrika*, **18**, 321.

Craig, C. C. (1929), "An application of Thiele's seminvariants to the sampling problem," *Metron*, **7**, 3.

Dressel, P. L. (1940), "Statistical seminvariants and their estimates with particular emphasis on their relation to algebraic invariants," *Ann. Math. Statist.*, **11**, 33.

Fisher, R. A. (1928), "Moments and product-moments of sampling distributions," *Proc. Lond. Math. Soc.*, **30**, 199–238.

—— (1930), "The moments of the distribution for normal samples of measures of departure from normality," *Proc. Roy. Soc.* A, **130**, 16.

—— and J. Wishart (1930), "The derivation of the pattern formulae of two-way partitions from those of simpler patterns," *Proc. Lond. Math. Soc.*, **32**, 195–208.

Hsu, C. T., and Lawley, D. N. (1939), "The derivation of the fifth and sixth moments of b_2 in samples from a normal population," *Biometrika*, **31**, 238.

Isserlis, L. (1931), "On the moment distributions of moments in the case of samples drawn from a limited universe," *Proc. Roy. Soc.*, **132** A, 586.

Kendall, M. G. (1940*a*), "Some properties of k-statistics," *Ann. Eugen. Lond.*, **10**, 106.

—— (1940*b*), "Proof of Fisher's rules for ascertaining the sampling semi-invariants of k-statistics," *Ann. Eugen. Lond.*, **10**, 215.

—— (1940*c*), "The derivation of multivariate sampling formulae from univariate formulae by symbolic operation," *Ann. Eugen. Lond.*, **10**, 392.

—— (1942), "On seminvariant statistics," *Ann. Eugen. Lond.*, **11**, 300.

Neyman, J. (1925), "Contributions to the theory of small samples drawn from a finite population," *Biometrika*, **17**, 472.

Pearson, E. S. (1930), "A further development of tests for normality," *Biometrika*, **22**, 239.

St. Georgescu, N. (1932), "Further contributions to the sampling problem," *Biometrika*, **24**, 65.

Thiele, T. N. (1903), *Theory of Observations*, London, C. and E. Layton (reprinted in *Ann. Math. Statist.*, **2**, 165).

Tschuprow, A. A. (1919), "On the mathematical expectation of the moments of frequency distributions," *Biometrika*, **12**, 140 and 185, and (1921) **13**, 283.

—— (1923), "On the mathematical expectation of moments of frequency distributions," *Metron*, **2**, 461 and 646.

Wishart, J. (1929*a*), "A problem in combinatorial analysis giving the distribution of certain moment statistics," *Proc. Lond. Math. Soc.*, **29**, 309.

—— (1929*b*), "The correlation between product moments of any order in samples from a normal population," *Proc. Roy. Soc. Edin.*, **49**, 78.

—— (1930), "The derivation of certain high-order sampling product-moments from a normal population," *Biometrika*, **22**, 224.

—— (1933), "A comparison of the semi-invariants of the distributions of moment and semi-invariant estimates in samples from an infinite population," *Biometrika*, **25**, 52.

EXERCISES

11.1. Show that the pattern functions of the following patterns :—

$$\begin{array}{ccc} \times & \times & \times \\ \times & . & \times \\ \times & \times & . \end{array} \qquad\qquad \begin{array}{cccc} \times & \times & \times & \times \\ . & . & \times & \times \\ \times & \times & . & . \end{array}$$

are $\dfrac{1}{(n-1)^2}$ and $\dfrac{}{n(n-1)^2}$ respectively.

(Fisher, 1928.)

11.2. Show that the pattern function of the pattern

$$\begin{array}{cccccc} \times & \times & \times & . & . & . \\ \times & \times & \times & . & . & . \end{array}$$

with p-columns is $$\frac{1}{n^{p-1}}\left\{1 - \frac{(-1)^{p-1}}{(n-1)^{p-1}}\right\}$$ (Fisher, 1928.)

11.3. Verify the formulæ of equations (11.33) and (11.39).

11.4. Show that the generating function of the moments of the k-statistics,

$$\Sigma\left\{\mu(p_1^{\pi_1} \ldots p_s^{\pi_s})\frac{t_{p_1}^{\pi_1}}{\pi_1!} \ldots \frac{t_{p_s}^{\pi_s}}{\pi_s!}\right.$$

is given by

$$\left[\exp\,\{t_1K_1 + t_2K_2 + t_3K_3 + \ldots\}\,\exp\left\{\kappa_1 s_1 + \frac{\kappa_2 s_2}{2!} + \ldots\right\}\right]_{x=0}$$

where K_r is the same function of the operators $\dfrac{\partial}{\partial x}$ as k_r is of the observations x and $s_r = \Sigma(x^r)$.

Deduce that

$$K_p(s_p) = p!$$
$$K_p(s_{p_1}s_{p_2} \ldots) = 0$$

where $(p_1p_2 \ldots)$ is any partition of p. (Fisher, 1928.)

Note that if $M(z)$ is the moment-generating function of z, that of $\xi = f(x)$ will be $\left[f\left(\dfrac{d}{dz}\right)M\right]_{z=0}$ and that of ξ^r by $\left[\left\{f\left(\dfrac{d}{dz}\right)\right\}^r M\right]_{z=0}$.
Hence that the generating function of the moments of ξ may be written

$$M(\zeta) = \left[\exp\left\{\zeta f\left(\frac{d}{dz}\right)\right\}M(z)\right]_{z=0}.$$

11.5. Show that

$$E\{k_p(x_1 + h, x_2, \ldots x_n)\} = \kappa_p + \frac{1}{n}h$$

and hence that

$$S_pk_p = p!$$
$$S_qk_p = 0 \qquad q \neq p$$

where S_p is the same function of the operators $\dfrac{\partial}{\partial x}$ as s_p is of the observations x.

(Kendall, 1940a.)

11.6. Show that the generating function of the moments of the k-statistics is given by

$$\left[\exp\left\{\frac{S_1\kappa_1}{1!} + \frac{S_2\kappa_2}{2!} + \ldots\right\}\exp\{k_1t_1 + k_2t_2 + \ldots\}\right],$$

and hence derive the result of the previous exercise.

(Kendall, 1941.)

11.7. Use Exercise 11.5 to show that, in the expression of k_p in terms of the symmetric sums s, the sum of the coefficients is $\frac{1}{n}$.

Show similarly that if

$$S_p = A_pk_p + A_{p-1,1}\,k_{p-1}k_1 + \ldots + A_{p_1p_2\ldots p_m}\,k_{p_1}k_{p_2}\ldots k_{p_m} + \ldots$$

then

$$1 = \frac{A_p}{n} + \frac{A_{p-1,1}}{n^2} + \ldots + \frac{A_{p_1\ldots p_m}}{n^m}$$

(Kendall, 1940*a*.)

11.8. Referring to the result of **11.22**, show that for a normal parent population the effect of adding a new part 2 to $\kappa(a_1^{\alpha_1}\ldots a_s^{\alpha_s})$ is to give pattern functions $\frac{1}{(n-1)}$ times those of the original. Show also that the effect on the numerical coefficient in an array is to multiply by twice the number of rows in the array. Deduce that the effect of adding a new part 2 is equivalent to operating by

$$\frac{2\kappa_2^2}{n-1}\,\frac{d}{d\kappa_2}.$$

(Fisher and Wishart, 1930.)

11.9. Use the previous exercise to establish equations (11.67) and (11.68).

11.10. In generalisation of Exercise 11.8 show that for a multivariate normal parent the effect of adding a covariance k_{pq} (p, q referring to the pth and qth variates) is equivalent to operating by

$$\sum_{rs}\frac{1}{(n-1)}(\kappa_{pr}\kappa_{qs} + \kappa_{ps}\kappa_{qr})\frac{d}{d\kappa_{rs}},$$

where κ_{pq} is the covariance of the variates p and q.

(Fisher and Wishart, 1930.)

11.11. If a pattern contains a column with three entries; if the patterns obtained by suppressing this column and (1) amalgamating the three rows, (2) amalgamating the pairs of three rows, and (3) leaving the rows unamalgamated are A, B_1, B_2, B_3 and C respectively, show that the pattern function for the original pattern is

$$\frac{n}{(n-1)(n-2)}A - \frac{1}{(n-1)(n-2)}(B_1 + B_2 + B_3) + \frac{2}{n(n-1)(n-2)}C.$$

Deduce that the function of the pattern

$$\begin{array}{ccc} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \cdot \end{array}$$

is

$$\frac{n^3 - 8n^2 + 17n + 2}{(n-1)^2(n-2)^2(n-3)}.$$

(Fisher and Wishart, 1930.)

11.12. Show that

$$\kappa\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{1}{n-1}\kappa_{11}^2 + \frac{1}{n-1}\kappa_{20}\kappa_{02}$$

and

$$\kappa\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{2}{n-1}\kappa_{11}^2$$

and hence that if

$$= \frac{\kappa_{11}}{\sqrt{(\kappa_{20}\kappa_{02})}} \text{ and } r = \frac{k_{11}}{\sqrt{(k_{20}k_{02})}},$$

then to order n^{-1} for normal samples

$$\text{var } r = \frac{1}{n}(1 - \rho^2)^2.$$

11.13. Show that for a bivariate normal population k_{tu} and k_{vw} have zero covariance unless $t + u = v + w$.

(Wishart, 1930.)

11.14. Show that for a bivariate normal population

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}$$

$$\text{var } (k_{tu}) = \kappa\begin{pmatrix} t & u \\ t & u \end{pmatrix} = t!u! \sum_{j=1}^{t+u} \frac{(j-1)!}{j!} \frac{\Delta^j 0^{t+u}}{n^{[j]}} \sigma_1^{2t}\sigma_2^{2u} F(-t, -u, 1, \rho^2),$$

where $\Delta^j 0^k$ is the jth difference of the kth power of zero and F refers to the hypergeometric function.

(Wishart, 1929*b*.)

11.15. Use the methods of this chapter to verify that to order n^{-1}

$$\text{var } (m_4) = \frac{1}{n}(\mu_8 - \mu_4^2 + 16\mu_3^2\mu_2 - 8\mu_5\mu_3).$$

11.16. Referring to Exercise 11.4, show that the moment-generating function $M'(\tau_2, \tau_3, \ldots)$ of the statistics $k_2, \dfrac{k_3}{k_2^{\frac{3}{2}}}, \dfrac{k_4}{k_2^2} \ldots \dfrac{k_r}{k_2^{\frac{r}{2}}} \ldots$

is given symbolically by

$$M'(\tau_2, \tau_3 \ldots) = \exp\left\{\tau_2 K_2 + \tau_3 \frac{K_3}{K_2^{\frac{3}{2}}} + \ldots\right\} M(t_2, t_3 \ldots),$$

where $M(t_2, t_3 \ldots)$ is the moment-generating function of the k-statistics and K_r is the same function of $\frac{\partial}{\partial t}$ as k_r is of x.

Noting that in normal samples the distribution of k_2 is independent of that of the other statistics and that its moment-generating function is $\left(1 - \frac{2\kappa_2 t_2}{n-1}\right)^{-\frac{1}{2}(n-1)}$ show that

$$M'(\tau_2, \tau_3 \ldots) = \exp\left(\tau_3 \frac{K_3}{K_2^{\frac{3}{2}}} + \ldots\right)\left(1 - \frac{2\kappa_2 t_2}{n-1}\right)^{-\frac{1}{2}(n-1)} M(t_3, t_4 \ldots).$$

Hence, if the number r refers to k_r, that

$$\mu(\ldots 5^c 4^b 3^a) = \mu(\ldots 5^c 4^b 3^a 2^{-j}) \frac{d^j}{dt^j}\left(1 - \frac{2\kappa_2 t}{n-1}\right)^{-\frac{1}{2}(n-1)}$$

where
$$2j = a + b + c + \ldots$$

and hence that $\mu(\ldots 5^c 4^b 3^a) = \mu(\ldots 5^c 4^b 3^a 2^{-j}) \frac{(n-1)(n+1)\ldots(n+2j-3)}{(n-1)^j}\kappa_2^j.$

Deduce that

$$\operatorname{var}\left(\frac{k_3}{k_2^{\frac{3}{2}}}\right) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)}$$

$$\mu_4\left(\frac{k_3}{k_2^{\frac{3}{2}}}\right) = \frac{108n^2(n-1)^2(n^2+27n-70)}{(n-2)^3(n+1)(n+3)(n+5)(n+7)(n+9)}.$$

Hence verify the formulae of equations (11.104) and (11.105). (This remarkable result is due to Fisher (1930). The independence of k_2 and the other statistics may be seen by considering the n-fold sample space, k_2 appearing as the square of a length and the others as angles (cf. Geary (1933), *Biometrika*, **25**, 184).)

11.17. Defining y by the relation

$$y = \sqrt{\frac{(n-1)(n-2)(n-3)}{24n(n+1)}} \frac{k_4}{k_2^2}$$

show that the moments of the distribution of y in samples from a normal population are

$$\mu_1 = 0$$

$$\mu_2 = 1 - \frac{12}{n} + \frac{88}{n^2} - \frac{532}{n^3} \ldots$$

$$\mu_3 = 6\sqrt{\frac{6}{n}}\left(1 - \frac{65}{2n} + \frac{4811}{8n^2} - \frac{136,605}{16n^3}\right).$$

$$\mu_4 = 3 - \frac{468}{n} + \frac{32,196}{n^2} - \frac{1,118,388}{n^3}$$

(E. S. Pearson, 1930, by the method of **11.23**, before the exact results of the previous exercise had been given. He fitted a Pearson Type IV to the distribution by using these moments, and tabulated the 1 per cent. and 5 per cent. significance points.)

CHAPTER 12

THE χ^2-DISTRIBUTION

12.1. Among the sampling distributions of current statistical theory the normal distribution is perhaps of widest application in virtue of the tendency of many statistics to normality in large samples, irrespective of the nature of the parent population. There is one other distribution, closely related to the normal, which has a somewhat similar general applicability and we give an account of it in this chapter.

Suppose we have a number of compartments or cells determined by specified ranges of a variate-value or by some qualitative character, such as the intervals of a univariate frequency-distribution, the cells of a bivariate distribution, or a simple classification of individuals into two classes, A and not-A. Suppose these cells are filled by random sampling from a parent population and that in the parent the proportion of members in the jth cell is π_j. In a sample of n there will occur proportions of, say, p_j in the jth cell, the observed numbers being accordingly np_j. If the sampling were such as to give an exact representation of the parent these numbers would be $n\pi_j$. Our fundamental problem is to determine how far the np_j can, to any acceptable degree of probability, diverge from the $n\pi_j$ by random sampling fluctuations. We shall then be able to test the accuracy of the hypothesis on which the π's were determined.

A few examples from material occurring in earlier chapters will illustrate the problem. In Table 5.1 on page 117 were given the actual occurrences of throws of dice, n being 26,306, the cells being eleven in number according to the number of "successes," and a third column showing the theoretical frequencies based on the hypothesis that the sampling obeyed a binomial law. The observed frequencies are our np's and the theoretical frequencies our $n\pi$'s. The question is, are the differences between the two such as can have arisen by sampling fluctuations alone? If not, then we must reject our hypothesis as to the generation of these dice-throws according to the binomial law.

Again, in the table of Example 8.6 were shown some results of inoculation against cholera. The question which interests us here is whether inoculation does in fact restrict or prevent attack. If it does not, we expect to find the same proportion of attacked in the inoculated class as in the not-inoculated class, e.g. the proportion of attacked in the former would be $\frac{69}{818} \times 279 = 23{\cdot}5$ approximately as against an observed 3. The former number is an $n\pi$ and the latter an np. Once again we should examine the differences, and if they were large enough to be inexplicable on the basis of sampling alone, should reject the hypothesis of independence of inoculation and attack, concluding that inoculation was to some extent preventive.

12.2. Consider then samples of n with a division of the possible classification into p cells; and suppose the members distributed simply at random in these cells.

Then the probability of there being l_1 members in the first cell, l_2 in the second, and so forth, is the term in $\pi_1^{l_1}\pi_2^{l_2}\ldots$ in the multinomial form

$$(\pi_1 + \pi_2 + \ldots \pi_p)^n,$$

that is to say, is

$$T = \pi_1^{l_1}\pi_2^{l_2} \ldots \pi_p^{l_p}\frac{n!}{l_1!\,l_2!\ldots l_p!} \quad . \quad . \quad . \quad . \quad (12.1)$$

If the l's are not small we have, in virtue of Stirling's approximation to the factorial,

$$T = \pi_1^{l_1} \dots \pi_p^{l_p} \frac{n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}}{l_1^{l_1+\frac{1}{2}}e^{-l_1}\sqrt{(2\pi)} \dots l_p^{l_p+\frac{1}{2}}e^{-l_p}\sqrt{(2\pi)}} \quad . \qquad (12.2)$$

and since $n = \Sigma l$ this becomes

$$T \propto \left(\frac{n\pi_1}{l_1}\right)^{l_1+\frac{1}{2}} \dots \left(\frac{n\pi_p}{l_p}\right)^{l_p+\frac{1}{2}} \qquad (12.3)$$

Now put

$$\lambda_j = n\pi_j$$

and

$$\xi_j = \frac{l_j - n\pi_j}{\sqrt{(n\pi_j)}} = \frac{l_j - \lambda_j}{\lambda_j^{\frac{1}{2}}}.$$

Then from (12.3) we have

$$\begin{aligned}\log T - \log (\text{constant}) &= \sum_j \left\{(l_j + \tfrac{1}{2}) \log \frac{\lambda_j}{l_j}\right\} \\ &= \Sigma(l_j + \tfrac{1}{2}) \log \frac{\lambda_j}{\lambda_j + \sqrt{\lambda_j}\xi_j} \\ &= -\Sigma(\lambda + \tfrac{1}{2} + \xi\sqrt{\lambda}) \log \left(1 + \frac{\xi}{\sqrt{\lambda}}\right).\end{aligned}$$

If λ is large, ξ will be small compared with λ and to first order we have, expanding the logarithm,

$$\begin{aligned}\log T - \log (\text{constant}) &= -\Sigma(\lambda + \tfrac{1}{2} + \xi\sqrt{\lambda})\left(\frac{\xi}{\sqrt{\lambda}} - \frac{1}{2}\frac{\xi^2}{\lambda}\right) \\ &= -\Sigma\{\tfrac{1}{2}\xi^2 + \xi\sqrt{\lambda} + 0(\lambda^{-\frac{1}{2}})\} \dots . \qquad (12.4)\end{aligned}$$

Now

$$\Sigma(\xi\sqrt{\lambda}) = \Sigma(l_j - \lambda_j) = n - n = 0$$

and hence, to order $\lambda^{-\frac{1}{2}}$,

$$\log T - \log \text{constant} = -\tfrac{1}{2}\Sigma\xi^2.$$

$$T \propto \exp(-\tfrac{1}{2}\Sigma\xi^2). \quad . \quad . \quad . \quad . \quad . \qquad (12.5)$$

Hence the frequency T varies as that of the sum of p normal variates of unit variance which are subject to the constraint $\Sigma(\xi\sqrt{\lambda}) = 0$ but otherwise independent.

Now put

$$\chi^2 = \Sigma\xi^2 = \Sigma\frac{(l_j - \lambda_j)^2}{\lambda_j}. \quad . \quad . \quad . \quad . \quad . \qquad (12.6)$$

Then the frequency of χ^2 is that of the sum of squares of $(p - 1)$ independent normal variates of unit variance. Its distribution is then, from Example 10.4, given by

$$dF = \frac{1}{2^{\frac{p-1}{2}}\Gamma\left(\frac{p-1}{2}\right)} e^{-\frac{1}{2}\chi^2}(\chi^2)^{\frac{1}{2}(p-3)}\, d(\chi^2) \qquad (12.7)$$

$$= \frac{1}{2^{\frac{1}{2}(p-3)}\Gamma\left(\frac{p-1}{2}\right)} e^{-\frac{1}{2}\chi^2}\chi^{p-2}\, d\chi. \quad . \qquad (12.8)$$

Furthermore, if there are certain constraints on the cell frequencies expressible by κ linear equations among them, the distribution remains of the same form but is now

$$dF = \frac{1}{2^{\frac{1}{2}(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{1}{2}\chi^2}\chi^{\nu-1}\, d\chi, \quad . \quad . \quad . \quad . \quad (12.9)$$

where ν, known as the number of degrees of freedom, is $\rho - \kappa$, that is the number of cells whose frequencies can be assigned without restriction. (Cf. Example 10.4.)

12.3. The distribution (12.9) is usually known as the χ^2-distribution, though it is actually that of χ. However, χ^2 and not χ is the quantity which always occurs in practical calculations and most tables of the distribution function have χ^2 as the argument. (12.9) is only an approximation, and relies on the fact that for large λ the Stirling approximation to the factorial will hold and that deviations from theoretical λ's are negligible to order $n^{-\frac{1}{2}}$. In point of fact, the approximation is very good and the χ^2-distribution may confidently be applied when the theoretical cell frequencies are, say, not less than 20.

Before dealing with the applications of the above results we will consider in more detail the properties of the distribution.

Properties of the χ^2-Distribution

12.4. Writing $\chi^2 = \zeta$ we have for the distribution

$$dF = \frac{1}{2^{\frac{1}{2}\nu}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{1}{2}\zeta}\zeta^{\frac{1}{2}\nu-1}\, d\zeta, \qquad 0 \leqslant \zeta \leqslant \infty \quad . \quad . \quad . \quad (12.10)$$

a Pearson Type III distribution. The characteristic function is

$$\phi(t) = \frac{1}{(1-2it)^{\frac{\nu}{2}}}; \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.11)$$

whence, for the cumulants, we have

$$\kappa_r = \nu\, 2^{r-1}(r-1)! \quad . \quad . \quad . \quad . \quad . \quad (12.12)$$

and for moments about the mean

$$\left.\begin{aligned} \mu_2 &= 2\nu \\ \mu_3 &= 8\nu \\ \mu_4 &= 48\nu + 12\nu^2 \\ \mu_5 &= 32\nu\,(5\nu + 12) \\ \mu_6 &= 40\nu(3\nu^2 + 52\nu + 96) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (12.13)$$

Since κ_r is linear in ν, μ_r, which can contain only $\left[\frac{r}{2}\right]$ powers of μ_2, must be of degree $\left[\frac{r}{2}\right]$ in ν, i.e. $\frac{r}{2}$ if r is even and $\frac{r-1}{2}$ if r is odd.

As ν tends to infinity the χ^2-distribution tends to normality, for in standard measure we have

$$\phi(t) = e^{-\frac{\nu it}{\sqrt{(2\nu)}}}\left(1 - \frac{2it}{\sqrt{(2\nu)}}\right)^{-\frac{\nu}{2}}$$

$$\log \phi(t) \rightarrow \frac{-\nu it}{\sqrt{(2\nu)}} - \frac{\nu}{2}\left\{\frac{-2it}{\sqrt{2\nu}} - \frac{1}{2}\left(\frac{2it}{\sqrt{2\nu}}\right)^2 \dots\right\}$$
$$\rightarrow -\tfrac{1}{2}t^2.$$

The tendency is, however, rather slow, and there are better approximations as we shall see in a moment.

12.5. The frequency curves given by (12.9) extend from zero to infinity. In the case $\nu = 1$ the curve is merely the positive half of the normal curve. In other cases it is zero at the origin, rises to a mode and then falls off again to infinity. The maximum ordinate of the χ-distribution (not the χ^2-distribution) is given by

$$\frac{d}{d\chi}(e^{-\frac{1}{2}\chi^2}\chi^{\nu-1}) = 0,$$

namely, by

$$\chi^2 = \nu - 1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.14)$$

and that of the χ^2-distribution or ζ-distribution by

$$\frac{d}{d\zeta}(e^{-\frac{1}{2}\zeta}\zeta^{\frac{1}{2}\nu-1}) = 0,$$

namely, by

$$\zeta = \chi^2 = \nu - 2. \quad . \quad (12.15)$$

The skewness of the ζ-distribution in the form (mean $-$ mode)/(standard deviation) is then

$$\frac{\nu - (\nu - 1)}{\sqrt{2\nu}} = \frac{1}{\sqrt{(2\nu)}}.$$

12.6. The distribution function of (12.10) is an incomplete Γ-function. We have

$$F(\zeta) = \int_0^{\zeta} \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{1}{2}\zeta}\zeta^{\frac{\nu}{2}-1}\,d\zeta$$
$$= \Gamma_{\zeta}\left(\frac{\nu}{2}\right),$$

or, in the notation of Pearson's tables,

$$= I\left(\frac{\zeta}{\sqrt{(\frac{1}{2}\nu + 1)}}, \tfrac{1}{2}\nu\right). \quad (12.16)$$

Some special tables have, however, been constructed.

(*a*) Elderton's table (*Tables for Statisticians and Biometricians*, Part I) gives the values of $P = \int_{\chi^2}^{\infty} dF = 1 - F(\zeta) = 1 - F(\chi^2)$ for values of $\nu = 2\,(1)\,29$ and $\chi^2 = 1\,(1)\,30$; $30\,(10)\,70$. In this table, which is to six places, our ν is denoted by $n' - 1$.

(*b*) A table by Yule, reproduced at the end of this volume, supplements Elderton's by giving P for $\nu = 1$, $\chi^2 = 0\,(0{\cdot}01)\,1$; $1\,(0{\cdot}1)\,10$.

(*c*) Kelley (1938) gives a four-place table of P for $\frac{\chi}{\sqrt{\nu}}$ from $0\,(0{\cdot}1)\,4{\cdot}1$ and $\nu = 1\,(1)\,10$; 12, 15, 19, 30.

(*d*) Fisher and Yates (1938) give tables in an inverted form, showing the values of

χ^2 for certain values of P and ν, namely $P = 0{\cdot}99$, $0{\cdot}98$, $0{\cdot}95$, $0{\cdot}90$ $(0{\cdot}10)$ $0{\cdot}10$, $0{\cdot}05$, $0{\cdot}02$, $0{\cdot}01$ and $0{\cdot}001$; and $\nu = 1\ (1)\ 30$.

(*e*) Thompson (*Biometrika*, **32** (1941), 187) gives tables in the inverted form for $P = 0{\cdot}995$, $0{\cdot}990$, $0{\cdot}975$, $0{\cdot}950$, $0{\cdot}750$, $0{\cdot}500$, $0{\cdot}250$, $0{\cdot}100$, $0{\cdot}050$, $0{\cdot}025$, $0{\cdot}010$, $0{\cdot}005$ and $\nu = 1\ (1)\ 30\ (10)\ 100$.

For general use the incomplete Γ-function tables are probably the best, as interpolation in Elderton's table does not give very great accuracy. The significance points tabled by Fisher and Yates are, however, sufficient for many practical applications of the χ^2-distribution in carrying out statistical tests. We reproduce at the end of the volume a diagram which will serve for such purposes. It shows, for co-ordinates ν and χ^2, the curves $P =$ constant, so that for given ν and χ^2 it is easy to determine whether P falls between any of the values for which the curves are drawn.

12.7. Except for Thompson's table, the tables do not cover the region for which $\nu > 30$, and for such values an approximation to the normal distribution may be employed. There are two such in common use:—

(*a*) (Fisher) that $\sqrt{(2\chi^2)}$ is normally distributed about mean $\sqrt{(2\nu - 1)}$ with unit variance;

(*b*) (Wilson and Hilferty (1931)) that $\left(\frac{\chi^2}{\nu}\right)^{\frac{1}{3}}$ is normally distributed about mean $1 - \frac{2}{9\nu}$ with variance $\sqrt{\frac{2}{9\nu}}$. The second is more accurate, but involves more arithmetical work in applications.

The relative speed of approach to normality of χ^2 and $\sqrt{(2\chi^2)}$ may be compared as follows:—

For χ^2 we have, from (12.13),

$$\gamma_1 = \sqrt{\beta_1} = \frac{8\nu}{(2\nu)^{\frac{3}{2}}} = \sqrt{\frac{2}{\nu}} \tag{12.17}$$

$$\gamma_2 = \beta_2 - 3 = \frac{12}{\nu} \tag{12.18}$$

For the moments of χ we have

$$\begin{aligned} \mu'_r &= \frac{1}{2^{\frac{1}{2}(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty e^{-\frac{\chi^2}{2}} \chi^{\nu+r-1}\, d\chi \\ &= 2^{\frac{r}{2}} \frac{\Gamma\left(\frac{\nu+r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \end{aligned} \tag{12.19}$$

Thus
$$\mu'_1 = \sqrt{2}\, \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$

Using the expansion

$$\log \Gamma(x+1) = \tfrac{1}{2}\log(2\pi) + (x + \tfrac{1}{2})\log x - x + \frac{1}{12x},$$

an extended form of Stirling's formula, we find after substitution and reduction

$$\mu_1' = \sqrt{\nu}\left(1 - \frac{1}{4\nu} + \frac{1}{32\nu^2} - \frac{17}{192\nu^3} + \ldots\right),$$

whence

$$\mu_1'^2 = \nu\left(1 - \frac{1}{2\nu} + \frac{1}{8\nu^2} - \frac{37}{192\nu^3} + \ldots\right)$$

Also

$$\begin{aligned} \mu_2' &= \nu \\ \mu_3' &= (\nu + 1)\mu_1' \\ \mu_4' &= (\nu + 2)\nu \end{aligned}$$

whence we find for moments about the mean

$$\mu_2 = \frac{1}{2} - \frac{-}{8\nu} + .$$

$$\mu_3 = \frac{}{4\sqrt{\nu}} + \ldots$$

$$\mu_4 = \frac{3}{4} + \frac{31}{48\nu}$$

Hence for the constants of $\sqrt{(2\chi^2)}$ we have

$$\mu_2 = 1 - \frac{1}{4\nu}$$

$$\gamma_1 = \frac{1}{\sqrt{2\nu}} + \ldots \qquad (12.20)$$

$$\gamma_2 = \frac{31}{12\nu} + \ldots \qquad (12.21)$$

A comparison with (12.17) and (12.18) shows that $\sqrt{(2\chi^2)}$ tends to normality with considerably greater rapidity than χ^2, γ_1 being only a half and γ_2 about a quarter of the values for the latter distribution. Moreover, the expression for μ_1' of χ is equal to $\sqrt{(\nu - \frac{1}{2})}$ to order $\nu^{-\frac{1}{2}}$ and hence $\sqrt{(2\chi^2)}$ is distributed about mean $\sqrt{(2\nu - 1)}$ to that order, with variance which is unity to order ν^{-1}.

12.8. For the Wilson-Hilferty approximation, consider the distribution of χ^2 about its mean value ν. Let us find the distribution of $\left(\frac{\chi^2}{\nu}\right)^h = y$, say, h as yet being undetermined. Write $\xi = \chi^2 - \nu$. Then

$$\begin{aligned} y &= \left(1 + \frac{\xi}{\nu}\right)^h \\ &= 1 + \frac{h\xi}{\nu} + \frac{h(h-1)}{2!}\frac{\xi^2}{\nu^2} + \text{etc.} \end{aligned}$$

Taking mean values and using the results of (12.13), we find, after some reduction,

$$\mu_1'(y) = 1 \quad \frac{h(h-1)\,\mu_2(\chi^2)}{} + \text{etc.}$$

$$= 1 \quad h(h-1)\ 1 \ - \frac{h(h-1)(h-2)(3h \quad 1)}{6\nu^2}$$

$$\frac{h^2(h-1)^2(h-2)(h-3)}{6\nu^3} + O(\nu^{-4}). \qquad (12.22)$$

TABLE 12.2

Distribution of First Hands at Whist according to Number of Cards held of a given Suit. Hands Dealt but not Played.

Number of Cards.	Observed Frequency. l	Theoretical Frequency. λ	$\frac{(l-\lambda)^2}{\lambda}$
0	35	43·5	1·661
1	290	272·2	1·164
2	696	700·0	0·023
3	937	973·5	1·369
4	851	811·3	1·943
5	444	424·0	0·943
6	115	141·3	4·895
7 and over . .	32	34·2	0·142
TOTAL	3400	3400·0	12·140

The total χ^2 is seen to be 12·140. The number of degrees of freedom ν is one fewer than the number of cells (excluding the total), namely 7. From the diagram in the appendix it is seen that the probability of getting a value as great as this or greater on random sampling $\left(=\int_{\chi^2}^{\infty} dF = P\right)$ lies between 0·1 and 0·05, very close to the former. The odds are therefore about 9 to 1 against getting the observed frequencies or frequencies which diverge to a greater extent from theoretical frequencies. This is hardly great enough for us to be able to say that an improbable event has occurred, and therefore we need not reject the hypothesis that the observed frequencies did in fact arise according to the hypergeometric law and that the discrepancies are merely sampling effects.

The matter stands differently with the distribution of Table 12.3 showing the distribution of 25,000 deals of whist classified in the same way as that of Table 12.2. Here $\chi^2 = 174{\cdot}130$

TABLE 12.3

Distribution of First Hands at Whist according to Number of Cards held of a given Suit. Hands actually Played.

Number of Cards.	Observed Frequency. l	Theoretical Frequency. λ	$\frac{(l-\lambda)^2}{\lambda}$
0	215	320	34·453
1	1,724	2,002	38·603
2	5,262	5,147	2·569
3	7,440	7,158	11·110
4	6,371	5,965	27·634
5	2,950	3,117	8·947
6	852	1,039	33·656
7	166	220	13·255
8 and over . .	20	31	3·903
TOTAL	25,000	24,999	174·130

and $\nu = 8$ (one more than in the previous example because we have grouped only those frequencies of 8 and over). From the diagram it is clear that the chance of getting such a value or a greater one is exceedingly small, certainly less than 1 in 10,000. This very rare event leads us to reject the hypothesis that the hypergeometric distribution is operating. The explanation is probably that these deals were taken from actual play, whereas those of the previous table were obtained without actually playing the hands. It is evident that in card play certain kinds of card (e.g. those of the same suit) tend to collect together and the shuffling is apt to be somewhat perfunctory. Thus the condition of realisation of the hypergeometric distribution, that the selection is random, was probably violated.

Example 12.2

In some classical experiments on pea-breeding Mendel obtained the following frequencies for different kinds of seeds in crosses from plants with round yellow seeds and wrinkled green seeds :—

	Observed	Theoretical
Round and yellow	315	312·75
Wrinkled and yellow	101	104·25
Round and green	108	104·25
Wrinkled and green	32	34·75
TOTAL	556	556·00

On the Mendelian theory of inheritance the frequencies should be in proportion 9, 3, 3, 1 and the theoretical frequencies are shown in the last column.

We find

$$\chi^2 = \frac{(2\cdot25)^2}{312\cdot75} + \frac{(3\cdot25)^2}{104\cdot25} + \frac{(3\cdot75)^2}{104\cdot25} + \frac{(2\cdot75)^2}{34\cdot75} = 0\cdot4700.$$

The number of degrees of freedom $\nu = 3$. The probability of obtaining the value of χ^2 or greater is seen to be between 0·9 and 0·95. There is thus nothing in the value of χ^2 to lead us to reject the Mendelian hypothesis.

12.11. Consider now a table of the type of Table 12.4, which shows the frequencies of a number of men according to eye colour and hair colour. If, on some hypothesis as to the relationship between eye and hair colour we determine theoretical frequencies in the body of the table, leaving the row and border columns unchanged, then there are a number of linear constraints on these frequencies.

In fact, if in such a table there are r rows and s columns it will be found that only $(r-1)(s-1)$ cells can be filled up arbitrarily. There are rs cells altogether ; but the fact that the rows and columns must add to assigned totals imposes $r+s$ constraints. These, however, are not independent, for the sum of the border column frequencies is equal to that of the border row frequencies and thus there are only $r+s-1$ independent linear constraints. Hence $rs-(r+s-1)=(r-1)(s-1)$ cells are independent and this is ν, the number of degrees of freedom associated with such a table.

Example 12.3

In Table 12.4 suppose that eye and hair colour are independent. Then the expected frequency in any cell with a row total x and a column total y will be $\frac{xy}{n}$, where n is the

TABLE 12.4

Distribution of 6800 Males according to Colour of Eye and Hair.
(Ammon, *Zur Anthropologie der Badener.*)

	Hair colour				
	Fair.	Brown.	Black.	Red.	TOTALS.
Blue	1768	807	189		2811
Grey or Green .	946	1387	746	53	3132
Brown .	115	438	288	16	857
TOTALS	2829	2632	1223	116	6800

total frequency. For instance, the expected number of men with fair hair and blue eyes is $\frac{2811 \times 2829}{6800} = 1169$. The theoretical frequencies obtained in this way are—

	Fair	Brown	Black	Red
Blue . . .	1169	1088	506	48·0
Grey or Green .	1303	1212	563	53·4
Brown . .	357	332	154	14·6

Hence
$$\chi^2 = \Sigma\frac{(1768-1169)^2}{1169}$$
$$= 1075{\cdot}2.$$
$$\nu = (4-1)(3-1) = 6.$$

The value of χ^2 is very improbable, P being less than 0·000,001. We accordingly reject the hypothesis of independence and conclude that hair colour and eye colour are associated.

12.12. It is useful to note that χ^2 may be put into a form which is sometimes more convenient for calculation. We have

$$\chi^2 = \Sigma\frac{(l-\lambda)^2}{\lambda}$$
$$= \Sigma\frac{l^2}{\lambda} - 2\Sigma l + \Sigma\lambda$$

n. . (12.28)

When the λ's are not integers it is easier to work with this formula, the squaring of the larger numbers l involving less arithmetic than the squaring of the smaller but non-integral numbers $l - \lambda$.

12.13. In the foregoing examples the theoretical frequencies λ were calculated without reference to the experimental data other than totals which merely subjected them to linear

constraints and hence preserved the Type III distribution of χ^2. There also arises the much more difficult case in which certain parameters necessary for the determination of the theoretical frequencies have to be determined from the data themselves. Suppose, for example, we attempt to represent a frequency-distribution by a normal curve. We have then to decide on the mean and variance of this curve, and they can, as a rule, only be estimated from the data themselves. The question then arises, what happens to the χ^2-distribution if, instead of the unknown parameters leading to the theoretical frequencies λ, we use estimates leading to the estimated theoretical frequencies, say, λ'? That is to say, how does the distribution of the statistic

$$\chi^2 = \Sigma\frac{(l - \lambda)^2}{\lambda}$$

compare with that of

$$\chi_1^2 = \Sigma\frac{(l - \lambda')^2}{\lambda'} \qquad . \quad (12.29)$$

This problem has not yet been completely solved. The nearest approach to a solution has been reached by R. A. Fisher (1924), who showed that χ_1^2 is distributed in the Type III form, provided that

(*a*) the sample is large and that each cell-frequency is large;

(*b*) that the number of degrees of freedom is reduced by unity for every parameter estimated;

(*c*) that the principle of estimation involved is such as to minimise χ^2. This is equivalent, for large samples, to taking a maximum likelihood estimate.

Departing from our usual practice, we shall have at this stage to state this result without proof. It cannot be adequately discussed until we have dealt with the principles of estimation in the second volume.

Example 12.4

The following table shows the distribution of 12 dice thrown 26,306 times, a 5 or 6 being reckoned a success. We have encountered these data before in Table 5.1.

TABLE 12.5

Distribution of 12 Dice thrown 26,306 times, a 5 or 6 being reckoned a Success.

Number of Successes.	Observed Frequency.	Frequency of $26,306(\frac{2}{3} + \frac{1}{3})^{12}$.	Frequency of $26,306(0{\cdot}6623 + 0{\cdot}3377)^{12}$.
0	185	203	187
1	1,149	1,217	1,146
2	3,265	3,345	3,215
3	5,475	5,576	5,465
4	6,114	6,273	6,269
5	5,194	5,018	5,115
6	3,067	2,927	3,043
7	1,331	1,254	1,330
8	403	392	424
9	105	87	96
10 and over	18	14	16
TOTALS	26,306	26,306	26,306

The third column shows the frequencies if the dice were perfect, that is the frequencies of the binomial law 26,306 $(\frac{2}{3} + \frac{1}{3})^{22}$. We find, in the usual way,

$$\chi^2 = \frac{(203 - 185)^2}{185} + \text{etc.} = 35{\cdot}941.$$

$$\nu = 10.$$

P is very small, less than 0·000,1 and we conclude that the hypothesis is to be rejected, i.e. that the dice were not perfect or that something was wrong with the sampling. In this particular case great care was in fact taken in rolling the dice and the balance lies in favour of rejecting the hypothesis that they were entirely unbiassed.

Let us then reconsider the data. If the dice are biassed, what is the true probability of getting a 5 or 6? This we must estimate from the data and it has already been seen that a maximum likelihood estimate is the mean number of successes in the sample itself. This is found to be 0·3377 and the last column in Table 12.5 shows the frequencies $26{,}306(0{\cdot}6623 + 0{\cdot}3377)^{12}$. The agreement with observation is evidently closer and we find now $\chi^2 = 8{\cdot}201$. ν is now 9, for we have estimated one parameter. P is now about 0·5, so that the observed frequencies are in quite good accord with theory.

12.14. Since χ^2 is the sum of a certain number of normal variates each with zero mean and unit variance, a number of different values of χ^2 may be added together and will be distributed in the Type III form with a number of degrees of freedom equal to the sum of the individual numbers. This result enables us to combine the results of a set of experiments so as to determine the probability of the whole set taken together. For example, Table 12.6 shows the data for inoculation against cholera on a certain tea estate.

TABLE 12.6

Inoculation against Cholera on a certain Tea Estate.

	Attacked.	Not Attacked.	Totals.
Inoculated	431		436
Not inoculated . . .	291		300
Totals	722	14	736

We find $\chi^2 = 3{\cdot}27$, $\nu = 1$ and P, from Appendix Table 7, about 0·071. This is small, but not small enough to reject the hypothesis, particularly when we note that the theoretical frequencies in the not-attacked column are far from large.

The results for six such estates were :—

χ^2	P	ν
3·27	0·071	1
9·34	0·0022	1
6·08	0·014	1
2·51	0·11	1
5·61	0·018	1
1·59	0·21	1
28·40		6

Here only one value of P is less than 0·01, and we might be inclined to doubt the reality

of association between inoculation and immunity. The sum of χ^2 is, however, 28·40, and $\nu = 6$, for which values we find $P < 0{\cdot}000{,}1$; so that together the results are significant.

The 2 × 2 Bivariate Table

12.15. We now return to a point which has been mentioned incidentally. If the theoretical frequencies in cells are small, the Type III distribution will hold only as an approximation depending on how closely the binomial distributions in individual cells are adequately represented by normal distributions. For some problems we can overcome the difficulty by grouping small frequencies, as has been done above in Example 12.1. But such a process sacrifices information and cannot always be carried out, e.g. in a 2 × 2 bivariate table.

Consider in the first place the symmetrical binomial $(\frac{1}{2} + \frac{1}{2})^{10}$. The second column in the following table shows the probability P that the number of successes in the first column will be at least attained (for the smaller of the pair) or attained or exceeded (for the larger of the pair).

Successes	P	P_1	P_2
0, 10	0·0010	0·0008	0·0022
1, 9	0·0108	0·0057	0·0134
2, 8	0·0547	0·0290	0·0569
3, 7	0·1719	0·1030	0·1714
4, 6	0·3770	0·2635	0·3759

If we regard this frequency as that of a single cell ($\nu = 1$) we should, for the corresponding χ^2-distribution, have the positive half of the normal curve. The values corresponding to $\chi^2 = \frac{1}{5}(5^2, 4^2, 3^2, 2^2, 1^2)$ are shown in the third column as P_1. They may be obtained from Appendix Tables 6 and 7, e.g. for the last term we have $\chi^2 = \frac{2}{5} = 0{\cdot}4$, $P = 0{\cdot}52709$, $P_1 = \frac{1}{2}$ of this $= 0{\cdot}2635$.

The correspondence between P and P_1 is evidently not very good. We can, however, improve it considerably by a correction due to Yates (1934). The distribution of χ^2 is continuous, whereas that of the binomial is not. To bring the two into comparability we really should consider the binomial frequency at the value r as spread over the range $r - \frac{1}{2}$ to $r + \frac{1}{2}$; or, what comes to the same thing, we should take χ^2 as half a unit less than the observed value in applying the Type III distribution. For example, for a deviation 3 corresponding to 8 successes we should take a deviation 2·5, giving $\chi^2 = \dfrac{2(2{\cdot}5)^2}{5} = 2{\cdot}5$ instead of 3. The values given by the corrected χ^2 are shown as P_2 above. The agreement between P_2 and P is evidently a great improvement on that between P_1 and P.

When the theoretical proportion in a cell is not $\frac{1}{2}$, the binomial distribution is skew, and there do not appear to be any simple corrections to compensate for this effect. The continuity correction will, however, result in an improvement if the theoretical frequency is near $\frac{1}{2}$ and is probably best made in all circumstances.

12.16. The 2 × 2 table may also be dealt with by exact methods. Consider in fact the table

$- d$

$b + d$

If the two variates are independent, the number of ways in which a table with such marginal totals can be constructed from the n sample members is

$$\binom{n}{a+c}\binom{n}{a+b} = \frac{n!}{(a+c)!\,(b+d)!}\,\frac{n!}{(a+b)!\,(c+d)!}.$$

The number of ways in which the body of the array can be completed is

$$\frac{n!}{a!\,b!\,c!\,d!}.$$

Consequently the probability of the distribution of the table is

$$\frac{(a+c)!\,(b+d)!\,(a+b)!\,(c+d)!}{n!\,a!\,b!\,c!\,d!}. \quad . \quad . \quad . \quad (12.30)$$

Thus the successive probabilities for $d = 0, 1, 2, \ldots$ are the successive terms in the hypergeometric series

$$F\{-(c+d),\ -(b+d),\ a-d+1,\ 1\}. \quad . \quad . \quad . \quad (12.31)$$

Example 12.5 (from F. Yates, 1934, quoting data by M. Hellman).

The following table shows the number of children classified according to the nature of the teeth and type of feeding.

	Normal Teeth.	Malocclluded Teeth.	TOTALS.
Breast fed		16	20
Bottle fed .		21	22
TOTALS			42

From (13.38) the probability of obtaining no normal breast-fed children if the attributes are independent is ($a = 0$)

$$\frac{5!\,37!\,20!\,22!}{42!\,20!\,0!\,5!\,17!} = 0{\cdot}03096.$$

The probabilities of obtaining 1, 2, . . . children are obtained by multiplying successively by $\frac{5 \times 20}{1 \times 18}$, $\frac{4 \times 19}{2 \times 19}$, $\frac{3 \times 18}{3 \times 20}$, and so on, and are as follows :—

Number of Normal Breast-fed Children.	Probability.
0	0·0310
1	0·1720
2	0·3440
3	0·3096
4	0·1253
5	0·0182
	1·0001

Thus the probability of getting four or more normal breast-fed children is 0·1435, and

we conclude that there is nothing to reject the hypothesis that breast feeding exerts no effect on the condition of the teeth. Had we used the χ^2 test in the ordinary way we should have found $P = 0{\cdot}0612$, less than half the true value. The continuity correction makes a great improvement, giving $P = 0{\cdot}1427$.

NOTES AND REFERENCES

The χ^2-distribution, though known to Helmert in 1876, was rediscovered and applied to statistical problems by Karl Pearson in 1900. In 1922 Yule and Fisher gave respectively experimental and theoretical evidence for what is now accepted as the correct method of determining the number of degrees of freedom in a bivariate table; but Pearson himself seems never to have acknowledged the soundness of this method, and some papers written between 1920 and 1930 on this subject are controversial and therefore not to be accepted uncritically.

For the use of the distribution in testing hypotheses when parent parameters have to be estimated, see Fisher (1924). For the exact test in a 2×2 table and the continuity correction, see Yates (1934).

More recently, Cochran (1936) and Haldane (1937*a*, 1937*b*, 1939*a*, 1939*b*) have discussed the distribution of χ^2 in bivariate tables when some hypothetical frequencies are small.

Cochran, W. G. (1936), "The χ^2-distribution for the binomial and Poisson series, with small expectations," *Ann. Eugen., Lond.*, **7**, 207.

—— (1938). "Note on J. B. S. Haldane's paper [1937*a* below]," *Biometrika*, **29**, 407.

Fisher, R. A. (1922), "On the interpretation of χ^2 from contingency tables and the calculation of P," *Jour. Roy. Statist. Soc.*, **85**, 87.

—— (1924), "The conditions under which χ^2 measures the discrepancy between observation and hypothesis," *Jour. Roy. Statist. Soc.*, **87**, 442.

Garwood, F. (1936), "Fiducial limits for the Poisson distribution," *Biometrika*, **28**, 437.

Haldane, J. B. S. (1937*a*), "The exact value of the moments of the distribution of χ^2 used as a test of goodness of fit when expectations are small," *Biometrika*, **29**, 133.

—— (1937*b*), "The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small," *Biometrika*, **29**, 389.

—— (1937*c*), "The approximate normalisation of a class of frequency distributions," *Biometrika*, **29**, 392.

—— (1939*a*), "Corrections to formulae in papers on the moments of χ^2," *Biometrika*, **31**, 220.

—— (1939*b*), "The mean and variance of χ^2 when used as a test of homogeneity when samples are small," *Biometrika*, **31**, 346.

—— (1939*c*), "The cumulants and moments of the binomial distribution and the cumulants of χ^2 for an $n \times 2$ fold table," *Biometrika*, **31**, 392.

Wilson, E. B., and Hilferty, M. M. (1931), "The distribution of chi-square," *Nat. Acad. Sci.*, **17**, 694.

Yates, F. (1934), "Contingency tables involving small numbers and the χ^2 test," *Supp. Jour. Roy. Statist, Soc.*, **1**, 217.

Yule, G. U. (1922), "An application of the χ^2 method to association and contingency tables with experimental illustrations," *Jour. Roy. Statist. Soc.*, **85**, 95.

EXERCISES

12.1. By the method of **12.8** show that

$$\left[\left(\frac{13\chi^2 - \nu}{12\nu}\right)^{\frac{1}{4}} + \frac{5}{18\nu}\left(1 + \frac{7}{48\nu}\right) - 1\right]\frac{6\sqrt{2\nu}}{5}\left(1 - \frac{1}{18\nu}\right)$$

is approximately normally distributed with unit variance about zero mean.

(Haldane, 1939.)

12.2. Use the χ^2-distribution to show that the distribution of digits from telephone directories (Table 1.4) could not in all probability have arisen by random sampling from a population in which each of the ten digits occurred with the same frequency.

12.3. Show that for a $2 \times n$ bivariate table $\nu = n - 1$ and

$$\chi^2 = \sum_j \frac{n_1 n_2\left(\frac{a_{1j}}{n_1} - \frac{a_{2j}}{n_2}\right)^2}{a_{1j} + a_{2j}}$$

where a_{1j}, a_{2j} are the frequencies in the jth column and n_1, n_2 are the border sums of the two rows.

12.4. Show that if ν is even

$$P = \int_\chi^\infty - e^{-\frac{\chi^2}{2}}\chi^{\nu-1}\,d\chi$$

$$= e^{-\frac{1}{2}\chi^2}\left(1 + \frac{\chi^2}{2} + \dots \frac{\chi^{\nu-2}}{2.4.6 \dots (\nu - 2)}\right)$$

and hence that the values of P for given χ^2 can be derived from tables of the Poisson exponential limit.

12.5. Show that in a 2×2 table whose frequencies are

b

$$\chi^2 = \frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(b + d)(a + c)}$$

the theoretical frequencies being those obtained on the hypothesis that the two variates are independent.

12.6. An experiment gives on hypothesis H $\chi^2 = 9$, $\nu = 8$. When repeated it gives the same result. Show that the two taken together do not give the same confidence in H as either taken separately.

12.7. (Data from *Report on the Spahlinger Experiments in Northern Ireland, 1931–1934*, H.M. Stationery Office, 1935.) In experiments on the immunisation of cattle from tuberculosis the following results were secured :—

	Died of Tuberculosis or very seriously affected.	Unaffected or only slightly affected.	TOTALS.
Inoculated with vaccine . . .	6	13	19
Not inoculated or inoculated with control media	8		11
TOTALS	14	16	30

Show that for this table, on the hypothesis that inoculation and susceptibility to tuberculosis are independent, $\chi^2 = 4{\cdot}75$, $P = 0{\cdot}029$; with a correction for continuity $P = 0{\cdot}039$; and that by the exact method of **12.15** $P = 0{\cdot}035$.

12.8. (Data from Yule, *Jour. Anthrop. Inst.*, 1906, **36**, 325.)

Sixteen pieces of photographic paper were printed down to different depths of colour from nearly white to a very deep blackish-brown. Small scraps were cut from each sheet and pasted on cards, two scraps on each card one above the other, combining scraps from the several sheets in all possible ways, so that there were 256 cards in the pack. Twenty observers then went through the pack independently, each one naming each tint either "light," "medium" or "dark."

The following table shows the name assigned to each of the two pieces of paper :—

Name assigned to Lower Tint.	Name assigned to Upper Tint.			TOTALS.
	Light.	Medium.	Dark.	
Light	850	571	580	2001
Medium	618	593	455	1666
Dark	540	456	457	1453
TOTALS	2008	1620	1492	5120

Show that there is a significant association between the name assigned to one piece and the name assigned to the other.

CHAPTER 13

ASSOCIATION AND CONTINGENCY

13.1. This and the next three chapters deal with the relationship between two or more variables. We shall consider populations, the members of which each bear one of each of several different sets of qualities or one value of each of several different variables, and shall discuss the measurement of the relationship among the qualities or variables in the populations. The corresponding questions of sampling will also fall for consideration. We may denote this branch of the theory by the general name of Theory of Dependence.

Association

13.2. Consider in the first instance a population classified according to whether each member bears or does not bear an attribute A. The presence of the attribute we may denote by A and the absence by α. We shall assume that each member must either be an A or an α, so that $\alpha =$ not-A and $A =$ not-α.

Suppose that each member of the population is classified according to two attributes A and B. Each may then be one of four kinds, AB, $A\beta$, αB and $\alpha\beta$. For example, if the attributes are the possession of blue eyes (A) and the possession of male sex (B), we shall have the four possible classes AB = blue-eyed males, $A\beta$ = blue-eyed females, αB = not-blue-eyed males, $\alpha\beta$ = not-blue-eyed females. Denoting the number in any class by the letters appropriate to that class in ordinary round brackets, we may then specify the population in the tabular form:—

	B's	not-B's	TOTALS.
A's	(AB)	$(A\beta)$	(A)
not-A's	(αB)	$(\alpha\beta)$	(α)
TOTALS	(B)	(β)	N

. (13.1)

or, more simply, by

a	b	$a + b$
c	d	$c + d$
$a + c$	$b + d$	N

. (13.2)

where $a = (AB)$, etc. Here N is the total number in the population.

13.3. If there is no relationship between the attribute A and the attribute B, that is to say if the possession of A is irrelevant to the possession of B, then there must be the same proportion of A's among the B's as among the β's. We thus *define* two attributes to be independent if

$$\frac{(AB)}{(B)} = \frac{(A\beta)}{(\beta)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (13.3)$$

or

$$\frac{a}{a+c} = \frac{b}{b+d} = \frac{a+b}{N}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (13.4)$$

It follows at once that each of the following is true :—

$$\frac{c}{a+c} = \frac{d}{b+d} = \frac{c+d}{N} \quad . \qquad (13.5)$$

$$\frac{a}{a+b} = \frac{c}{c+d} = \frac{a+c}{N} \quad . \qquad (13.6)$$

$$\frac{b}{a+b} = \frac{d}{c+d} = \frac{b+d}{N} \quad . \qquad (13.7)$$

These are derivable by simple algebra from (13.4) and, in words, will be found to express the same fundamental fact that the proportion of members bearing an attribute X is the same among the Y's as among the not-Y's. It also follows that

$$a = \frac{(a+b)(a+c)}{N} \quad . \qquad (13.8)$$

with three similar equations in b, c and d.

If now in any given table

$$a > \frac{(a+b)(a+c)}{N} \quad . \qquad (13.9)$$

or, in the alternative notation,

$$(AB) > \frac{(A)(B)}{N} \quad . \qquad (13.10)$$

we shall say that A and B are positively associated. *Per contra*, if

$$(AB) < \frac{(A)(B)}{N} \quad . \quad . \quad . \quad . \qquad (13.11)$$

they are said to be negatively associated or disassociated.

Example 13.1

Association between inoculation against cholera and exemption from attack. (Greenwood and Yule (1915), *Proc. Roy. Soc. Medicine*, **8**, 113.) The following table shows 818

	Not Attacked.	Attacked.	Totals.
Inoculated			279
Not-inoculated . . .	473		539
Totals	749	69	818

cases classified according to inoculation against cholera (attribute A) and freedom from attack (attribute B).

If the attributes were independent the frequency in the inoculated-not-attacked class would be $\frac{279 \times 749}{818} = 255$. The observed frequency is greater than this and hence inoculation is positively associated with exemption from attack.

13.4. The reader will recognise in this example a type of 2×2 table which was discussed in connection with the χ^2-distribution. In fact, if the data are considered as a sample there arises at once the question how far the positive association, which certainly exists in the sample, is indicative of real association in the parent population. The χ^2-distribution, as shown in the previous chapter, provides an objective method of forming a judgment on this matter. χ^2 itself, however, does not provide an adequate measure of the intensity of the association. Altogether apart from sampling questions, we sometimes wish to compare the strength of associations in different populations or between different attributes, and some coefficients proposed for the purpose will now be considered.

13.5. The more obvious desiderata in a coefficient-measuring association are (a) that it shall vanish when the attributes are independent; (b) that it shall be $+1$ when there is complete positive association and -1 when there is complete negative association; (c) that it should increase as the frequencies proceed from dissociation to association. As to this latter point, consider the difference between observed and "independence" frequencies in the cell corresponding to (AB), viz.:—

$$\delta = (AB) - \frac{(A)(B)}{N}. \qquad (13.12)$$

Since the border frequencies are constant it is evident that the difference in any cell between observed and "independence" frequencies is $\pm\delta$ and thus δ determines uniquely the departure from independence. We may interpret condition (c) as meaning that our coefficient should increase with δ. It may be noted that

$$\delta = a - \frac{(a+b)(a+c)}{a+b+c+d}$$
$$= \frac{ad - bc}{N}. \qquad (13.13)$$

Following Yule (1900, 1912) we define the coefficient of association Q by the equation

$$Q = \frac{ad - bc}{ad + bc} = \frac{N\delta}{ad + bc}. \qquad (13.14)$$

It is zero if the attributes are independent, for then $\delta = 0$. It can equal $+1$ only if $bc = 0$, in which case there is complete association (either no α's are B's or no A's are β's), and -1 only if $ad = 0$, in which case there is complete disassociation. Furthermore, Q increases with δ, for if $\varepsilon = \frac{bc}{ad}$

then
$$Q = \frac{1 - \varepsilon}{1 + \varepsilon}$$

and $\frac{dQ}{d\varepsilon}$ is negative, as is $\frac{d\delta}{d\varepsilon}$, so that $\frac{dQ}{d\delta}$ is positive.

A somewhat similar coefficient, also due to Yule, is the coefficient of colligation

$$Y = \frac{1 - \sqrt{\frac{bc}{ad}}}{1 + \sqrt{\frac{bc}{ad}}} \quad . \quad . \quad . \quad . \quad . \quad (13.15)$$

This also satisfies our conditions, as the reader may verify for himself.

13.6. Yet a third coefficient, which will be shown below to be related to χ^2, is

$$V = \frac{N\delta}{+\sqrt{(A)(\alpha)(B)(\beta)}}$$

$$= \frac{(ad - bc)}{+\{(a + b)(a + c)(b + d)(c + d)\}^{\frac{1}{2}}} \quad . \quad (13.16)$$

This is evidently zero when $\delta = 0$ and increases with δ. If $V = 1$ we have

$$(a + b)(a + c)(b + d)(c + d) = (ad - bc)^2,$$

giving

$$4abcd + a^2(bc + bd + cd) + b^2(ac + ad + cd) + c^2(ab + ad + bd) + d^2(ac + ab + bc) = 0.$$

Since no frequency can be negative this can only vanish if at least two of a, b, c, d are zero. If the frequencies in the same row and column vanish the case is purely nugatory. We have then only to consider $a = 0, d = 0$ or $b = 0, c = 0$. In the first case $V = 1$, in the second $V = -1$. It cannot lie outside these limits.

13.7. It will be observed that whereas V is unity only if two frequencies in the 2×2 table vanish, Q and Y are unity if only one frequency vanishes. This raises a point in connection with the definition of *complete* association. We shall say that association is complete if all A's are B's, notwithstanding that all B's are not A's. If all dumb men are deaf there is complete association between dumbness and deafness, however many deaf men there are who are not dumb. The coefficient V is unity only if all A's are B's and all B's are A's, a condition which we could, if so desired, describe as *absolute* association.

It is necessary to point out in this connection that statistical association is different from association in the colloquial sense. In current speech we say that A and B are associated if they occur together fairly often; but in statistics they are associated only if A occurs relatively more or less frequently among the B's than among the not-B's. If 90 per cent. of smokers have poor digestions we cannot say that smoking and poor digestion are associated until it is shown that less than 90 per cent. of non-smokers have poor digestions.

Example 13.2

Consider again the data of Example 13.1. For the various coefficients we have

$$Q = \frac{(276 \times 66) - (3 \times 473)}{(276 \times 66) + (3 \times 473)} = 0{\cdot}8555$$

$$Y = \frac{1 - \sqrt{\dfrac{3 \times 473}{276 \times 66}}}{1 + \sqrt{\dfrac{3 \times 473}{276 \times 66}}} = 0{\cdot}5636$$

$$V = \frac{(276 \times 66) - (3 \times 473)}{\sqrt{(279 \times 539 \times 749 \times 69)}} = 0{\cdot}1905.$$

These values are, as might be expected, different, although they all refer to the same intensity of association. Comparisons, however, naturally fall to be made between values of coefficients of the same type, and the fact that different types give different values does not affect their usefulness or the comparability of members of any one type.

13.8. The methods of Chapter 9 may be used to give the standard errors of the three coefficients based on material obtained by random sampling.

We recall that for any of the four frequencies a, b, c, d we have results such as

$$\text{var } a = \frac{a(N - a)}{N} \qquad (13.17)$$

$$\text{cov } (a, b) = -\frac{ab}{N}. \qquad (13.18)$$

The first is merely another way of writing the expression for the variance of a binomial. The second follows from

$$0 = \text{var } (a + b) = \text{var } a + \text{var } b + 2 \text{ cov } (a, b).$$

Then for $\varepsilon = \dfrac{bc}{ad}$ we have, writing Δ for the differential to avoid confusion with the frequency d,

$$\frac{\Delta\varepsilon}{\varepsilon} = \frac{\Delta b}{b} + \frac{\Delta c}{c} - \frac{\Delta a}{a} - \frac{\Delta b}{b},$$

whence

$$\frac{\text{var } \varepsilon}{\varepsilon^2} = \Sigma \frac{\text{var } a}{a^2} + 2\, \Sigma \left\{ \pm \frac{\text{cov}(a, b)}{ab} \right.$$

Substitution from (13.17) and (13.18) gives

$$\text{var } \varepsilon = \varepsilon^2 \left(\frac{1}{a} + \frac{1}{b} \right. \qquad (13.19)$$

Now

$$Q = \frac{1 - \varepsilon}{1 + \varepsilon}$$

and hence

$$\frac{\Delta Q}{Q} = -\frac{2\, \Delta\, \varepsilon}{1 - \varepsilon^2},$$

giving

$$\text{var } Q = \frac{4Q^2 \text{ var } \varepsilon}{(1 - \varepsilon^2)^2}$$

$$= \frac{(1 - Q^2)^2}{4} \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right). \qquad (13.20)$$

In a similar way we have

$$\text{var } Y = \frac{(1 - Y^2)^2}{16} \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right). \qquad (13.21)$$

The sampling variance of V may be found similarly but involves rather more lengthy algebra. Yule (1912) gives the result

$$\operatorname{var} V = \frac{1}{N}\left[1 - V^2 + \left(V + \tfrac{1}{2}V^3\right)\frac{(a-d)^2 - (b-c)^2}{\{(a+b)(a+c)(b+d)(c+d)\}^{\frac{1}{2}}}\right.$$

$$\left. - \tfrac{3}{4}V^2\left\{\frac{(a+b-c-d)^2}{(a+b)(c+d)} + \frac{(a+c-b-d)^2}{(a+c)(b+d)}\right\}\right] . \qquad (13.22)$$

In applying these formulae it is, as usual in large-sample theory, assumed that the observed frequencies may be used instead of theoretical frequencies in the sampling variances.

Example 13.3

Reverting to Example 13.2, we have for the standard error of Q

$$\frac{1 - Q^2}{2}\sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}$$

$$\frac{(0{\cdot}8555)}{2}\sqrt{\left(\frac{1}{276} + \frac{1}{3} + \frac{1}{473} + \frac{1}{66}\right)}$$

$$= 0{\cdot}0798.$$

The coefficient Q thus probably lies in the range $0{\cdot}856 \pm 0{\cdot}239$ in the population from which these data were derived, assuming of course that the sampling was random.

Partial Association

13.9. The coefficients described above measure the dependence of two attributes in the statistical sense, but in order to decide whether such dependence has any causal significance it is often necessary to consider associations in sub-populations. Suppose, for example, a positive association is noticed between inoculation and exemption from attack. It is natural to infer that the inoculation confers exemption, but this is not necessarily so. It might be that the people who are inoculated are drawn largely from the richer classes, who live in better hygienic conditions and are therefore better equipped to resist attack or less exposed to risk. In other words, the association of A and B might be due to the association of both with a third attribute C (wealth).

Now it is clear that this explanation would not hold if the hygienic circumstances were constant in the population. If we then consider the association of A and B in the sub-populations (C) (well-to-do classes) and (γ) (poorer classes) and find that it persists, the explanation is rejected. Furthermore, if the association in (γ) was weaker than that in (C), there would be some indication that hygienic conditions are related to exemption from attack, though not constituting the only factor concerned.

13.10. Associations in sub-populations are called partial associations. Analogously to (13.9), A and B are said to be positively associated in the population of C's if

$$(ABC) > \frac{(AC)(BC)}{(C)} . \qquad (13.23)$$

where (ABC) represents the number of members bearing the attributes A, B and C; and so on. We may also define coefficients of partial association, colligation, etc., such as

$$Q_{AB.C} = \frac{(ABC)(\alpha\beta C) - (A\beta C)(\alpha BC)}{(ABC)(\alpha\beta C) + (A\beta C)(\alpha BC)}, \qquad (13.24)$$

which is derived from (13.14) by adding C to all the symbols representing the frequencies.

Example 13.4

Galton's " Natural Inheritance " gives some particulars, for 78 families containing not less than six brothers or sisters, of eye-colour in parent and child. Denoting a light-eyed child by A, a light-eyed parent by B and a light-eyed grandparent by C, we trace every possible line of descent and record whether a light-eyed child has light-eyed parent and grandparent, the number of such being denoted by (ABC) and so on. The symbol $(A\beta\gamma)$, for example, denotes the number of light-eyed children whose parents and grandparents have not light eyes. The eight possible classes are

$(ABC) = 1928$	$(\alpha BC) = 303$
$(AB\gamma) = 596$	$(\alpha B\gamma) = 225$
$(A\beta C) = 552$	$(\alpha\beta C) = 395$
$(A\beta\gamma) = 508$	$(\alpha\beta\gamma) = 501$

The first question we discuss is: does there exist any association between parent and offspring with regard to eye-colour? We consider both the grandparent-parent group (association of B's and C's) and the parent-child group (association of A's and B's). We have, for the former:

Proportion of light-eyed among children of light-eyed parents,

$$\frac{(BC)}{(C)} = \frac{2231}{3178} = 70{\cdot}2 \text{ per cent.}$$

Proportion of light-eyed among children of not-light-eyed parents,

$$\frac{(B\gamma)}{(\gamma)} = \frac{821}{1830} = 44{\cdot}9 \text{ per cent.};$$

and for the latter, analogously,

$$\frac{(AB)}{(B)} = \frac{2524}{3052} = 82{\cdot}7 \text{ per cent.}$$

$$\frac{(A\beta)}{(\beta)} = \frac{1060}{1956} = 54{\cdot}2 \text{ per cent.}$$

Frequencies such as $(A\beta)$ are calculable direct from the eight classes given above, e.g.

$$(A\beta) = (A\beta C) + (A\beta\gamma) = 552 + 508 = 1060.$$

Evidently there is some positive association between parent and offspring in regard to eye-colour.

Consider now the relationship between eye-colours of grandparents and grandchildren. We have:

Proportion of light-eyed among grandchildren of light-eyed grandparents

$$= \frac{(AC)}{(C)} = \frac{2480}{3178} = 78{\cdot}0 \text{ per cent.}$$

Proportion of light-eyed among grandchildren of not-light-eyed grandparents

$$= \frac{(A\gamma)}{(\gamma)} = \frac{1104}{1830} = 60{\cdot}3 \text{ per cent.}$$

The association between grandparents and grandchildren is also positive.

In tabular form the data are :—

Grandparents.

	C	γ	TOTALS.
B	2231	821	3052
β	947	1009	1956
TOTALS	3178	1830	5008

Parents.

Chi

	B	β	TOTALS.
A	2524	1060	3584
α	528	896	1424
TOTALS	3052	1956	5008

Grandparents.

Gi

	C	γ	TOTALS.
A	2480	1104	3584
α	698	726	1424
TOTALS	3178	1830	5008

The coefficients of association Q and Y are

	Q	Y
Grandparents—parents	0·487	0·260
Parents—children	0·603	0·336
Grandparents—grandchildren . . .	0·401	0·209

Now the question arises : is the resemblance between grandparent and grandchild due merely to that between grandparent and parent, parent and child ? To investigate this, we must consider the associations of grandparent and grandchild in the sub-populations "parents light-eyed" and "parents not-light-eyed," that is, the associations of A and C in B and β. We have :—

Parents Light-eyed

Proportion of light-eyed amongst grandchildren of light-eyed grandparents

$$= \frac{(ABC)}{(BC)} = \frac{1928}{2231} = 86{\cdot}4 \text{ per cent.}$$

Proportion of light-eyed amongst grandchildren of not-light-eyed grandparents

$$= \frac{(AB\gamma)}{(B\gamma)} = \frac{596}{821} = 72 \cdot 6 \text{ per cent.}$$

Parents not Light-eyed

Proportion of light-eyed amongst the grandchildren of light-eyed grandparents

$$= \frac{(A\beta C)}{(\beta C)} = \frac{552}{947} = 58 \cdot 3 \text{ per cent.}$$

Proportion of light-eyed amongst the grandchildren of not-light-eyed grandparents

$$= \frac{(A\beta\gamma)}{(\beta\gamma)} = \frac{508}{1009} = 50 \cdot 3 \text{ per cent.}$$

In both cases the partial association is well marked and positive. The association between grandparents and grandchildren cannot, then, be due wholly to the associations between grandparents and parents, parents and children. There is ancestral heredity, as it is called, as well as parental heredity. The relevant four-fold tables are :—

Parents light-eyed
Grandparents.

	BC	$B\gamma$	TOTALS.
AB	1928	596	2524
αB	303	225	528
TOTALS	2231	821	3052

Parents not-light-eyed.
Grandparents.

	βC	$\beta\gamma$	TOTALS.
$A\beta$	552	508	1060
$\alpha\beta$	395	501	896
TOTALS	947	1009	1956

The coefficients of association and colligation are :—

$$Q_{AC.B} = 0 \cdot 412 \qquad Q_{AC.\beta} = 0 \cdot 159$$
$$Y_{AC.B} = 0 \cdot 216 \qquad Y_{AC.\beta} = 0 \cdot 080$$

13.11. If there are p different attributes under consideration the number of partial associations can become very large, even for moderate p. For example, we can choose two in $\binom{p}{2}$ ways and consider their associations in all the possible sub-populations of the other $(p-2)$, which are seen to be 3^{p-2} in number. Thus there are $\binom{p}{2}3^{p-2}$ associations. In practice, however, we need only consider a few of them.

One result in this connection is worth noticing. We have, generalising δ of equation (13.13) :—

$$\delta_{AB.C} + \delta_{AB.\gamma} = \left\{(ABC) - \frac{(AC)(BC)}{(C)}\right\} + \left\{(AB\gamma) - \frac{(A\gamma)(B\gamma)}{(\gamma)}\right\}$$

$$= (AB) - \frac{(A)(B)}{N} - \frac{N}{(C)(\gamma)}\left\{(AC)(BC) - \frac{(A)(C)(BC)}{N} - \frac{(B)(C)(AC)}{N} + \frac{(A)(B)(C)^2}{N}\right\}$$

$$= \delta_{AB} - \frac{N}{(C)(\gamma)}\left\{(AC) - \frac{(A)(C)}{N}\right\}\left\{(BC) - \frac{(B)(C)}{N}\right\}$$

$$= \delta_{AB} - \frac{N}{(C)(\gamma)}\delta_{AC}\,\delta_{BC}. \qquad (13.25)$$

If, then, A and B are independent in both (C) and (γ), $\delta_{AB.C} = \delta_{AB.\gamma} = 0$ and

$$\delta_{AB} = \frac{N}{(C)(\gamma)}\delta_{AC}\,\delta_{BC}, \qquad (13.26)$$

i.e. they are not independent in the population as a whole unless C is independent of A or B or both in that population.

This peculiar result indicates that illusory associations may arise when two populations (C) and (γ) are amalgamated, or that real associations may be obscured. If A and C, B and C, are associated we have, from (13.25),

$$\delta_{AB} = \frac{N}{(C)(\gamma)}\,\delta_{AB}\,\delta_{AC} + \delta_{AB.C} + \delta_{AB.\gamma}$$

so that even if A and B are independent in (C) and (γ) they will appear as associated in the two together. Again, if A and B are associated positively in (C) and negatively in (γ), δ_{AB} may be zero, that is to say, they may appear as independent in the whole population.

Example 13.5

Consider the case in which a number of patients are treated for a disease and there is noted the number of recoveries. Denoting A by recovery, α by not-recovery, B by treatment, β by not-treatment, suppose the frequencies are

	B	β	TOTALS.
A	100	200	300
α	50	100	150
TOTALS	150	300	450

Here $(AB) = 100 = \frac{(A)(B)}{N}$, so that the attributes are independent. So far as can be seen, treatment exerts no effect on recovery.

Denoting male sex by C and female sex by γ, suppose the frequencies among males and females are

Males

	BC	βC	Totals.
AC	80	100	180
αC	40	80	120
Totals	120	180	300

Females

	$B\gamma$	$\beta\gamma$	Totals.
$A\gamma$	20	100	120
$\alpha\gamma$	10	20	30
Totals	30	120	150

In the male group we now have

$$Q_{AB.C} = \frac{(80 \times 80) - (100 \times 40)}{(80 \times 80) + (100 \times 40)} = 0{\cdot}231$$

and in the female group

$$Q_{AB.\gamma} = -\ 0{\cdot}429.$$

Thus among the males treatment is positively associated with recovery, and among the females negatively associated. The apparent independence in the two together is due to the cancelling of these associations in the sub-populations.

Contingency

13.12. We now turn to the more general case in which a population is divided into a number of categories $A_1, A_2 \ldots A_p$ instead of simply dichotomised into two, A and not-A. If there is a second classification into $B_1, B_2 \ldots B_q$ the frequencies may be arranged in the form :—

	A_1	A_2	. . .	A_p	Totals.
B_1	(A_1B_1)	(A_2B_1)	. . .	(A_pB_1)	(B_1)
B_2	(A_1B_2)	(A_2B_2)	. . .	(A_pB_2)	(B_2)
. . .	. . .	. . .	. . .	. . .	. . .
B_q	(A_1B_q)	(A_2B_q)	. . .	(A_pB_q)	(B_q)
Totals	(A_1)	(A_2)	. . .	(A_q)	N

. . (13.27)

This is known as a contingency table. We have already encountered an example in Table 12.4. Ordinary bivariate frequency tables can, of course, be regarded as contingency tables, but there is a difference: in the bivariate table the order of rows and columns is determined by the variate-values, whereas in the contingency table the order of rows and columns is, in general, arbitrary.

In (13.27) the frequency in the ith column and jth row is denoted by (A_iB_j). As in the case of the 2×2 table we write

$$\delta_{ij} = (A_iB_j) - \frac{(A_i)(B_j)}{N} \quad . \quad . \quad . \qquad (13.28)$$

and define the attributes as independent if every δ is zero. We have:

$$\begin{aligned} \sum_{i,\,j} \delta_{ij} &= \Sigma\left\{(A_iB_j) - \frac{(A_i)(B_j)}{N}\right. \\ &= N - \frac{N^2}{N} = 0 \end{aligned} \qquad (13.29)$$

and, in conformity with the notation of (12.6) we have

$$\chi^2 = \Sigma \frac{N\delta^2_{ij}}{(A_i)(B_j)}. \quad . \quad . \quad . \quad . \quad . \qquad (13.30)$$

χ^2 is sometimes called the square contingency and $\frac{\chi^2}{N}$ the mean square contingency.

13.13. We have already seen in Chapter 12 how χ^2 may be used to test the hypothesis that the observed frequencies could have arisen by random sampling from a population in which the attributes were independent. We now proceed to consider the construction of measures of dependence.

Evidently $\chi^2 = 0$ if and only if each $\delta = 0$. Thus χ^2 vanishes if and only if the attributes are independent in the observed population. Furthermore, as χ^2 becomes larger the observed frequencies deviate more and more from the "independence" frequencies; it thus provides some sort of measure of the strength of relationship in contingency tables. For example, in the 2×2 table we have

$$V(\text{equation (13.16)}) = \frac{\chi^2}{N}, \qquad (13.31)$$

which illustrates the relationship between V and χ^2.

χ^2 itself, however, does not constitute a very useful coefficient since it may increase without limit. Following Karl Pearson we may put

$$= \sqrt{\overline{N}} \quad {}^{2} \quad . \quad . \quad . \quad . \quad . \qquad (13.32)$$

and call C Pearson's coefficient of contingency. Even this has its limitations. It vanishes, as it should, when there is complete independence; but in general it cannot attain unity. Consider, for example, a $t \times t$ table in which the diagonals (A_iB_i) are of frequency α_i and

all other compartments are zero. Obviously no greater degree of dependence is possible. We then have

$$\delta_{ii} = \alpha_i - \frac{\alpha_i^2}{N}$$

$$\frac{N\Sigma\left(\alpha_i - \frac{\alpha_i^2}{N}\right)^2}{\alpha_i^2}$$

$$= (N - 1)t$$

and

$$= \sqrt{\frac{t-1}{t}}. \qquad (13.33)$$

If $t = 5$, for example, the maximum value of C is 0.894.

To remedy this effect Tschuprow proposed the coefficient

$$T = \left\{\frac{\chi^2}{N\sqrt{(p-1)(q-1)}}\right\}^{\frac{1}{2}}. \qquad (13.34)$$

This can attain unity when $p = q$; but it is still not clear how it behaves when p and q are unequal.

Example 13.6

TABLE 13.1

Distribution of Schoolchildren according to Intelligence and Standard of Clothing.
(From W. H. Gilby (1911), *Biometrika*, **8**, 94.)

	A and *B*	*C*	*D*	*E*	*F*	*G*	TOTALS.
Very well clad . . .	33	48	113	209	194	39	636
Well clad	41	100	202	255	138	15	751
Poor but passable . .	39	58	70	61	33	4	265
Very badly clad . .	17	13	22	10	10	1	73
TOTALS	130	219	407	535	375	59	1725

The above table shows the distribution of 1725 schoolchildren who were classified (1) according to their standard of clothing and (2) according to their intelligence, the standards in the latter case being A = mentally deficient, B = slow and dull, C = dull, D = slow but intelligent, E = fairly intelligent, F = distinctly capable, G = very able. Required to discuss whether there is any association between standards of clothing and intelligence.

We note in the first place that a table of this kind could, theoretically, be discussed

by considering all the possible 2×2 comparisons to be extracted from it; e.g. for the corners of the table we have

	A and B		TOTALS.
Very well clad . . .	33	39	72
Very badly clad . . .	17		18
TOTALS	50	40	90

Here, for example, 54 per cent. of the very well clad were very able, but only 5 per cent. of the very badly clad. However, what we really require is not a series of individual comparisons of this kind but a general comparison over the whole table, and it is for such purposes that the coefficient of contingency is designed.

We then proceed to work out the "independence" frequencies, e.g. that in the top left-hand corner of the table is $\frac{636 \times 130}{1725} = 47{\cdot}930$. The contribution to χ^2 from this compartment is then $\frac{(14{\cdot}930)^2}{47{\cdot}930} = 4{\cdot}651$. It will be found that the sum of the contributions from the 20 compartments is 174·92. We then have

$$C = \sqrt{\frac{174{\cdot}92}{1725 + 174{\cdot}92}} = 0{\cdot}303,$$

indicating a considerable degree of association. For the Tschuprow coefficient we have

$$T = \sqrt{\frac{174{\cdot}92}{1725\sqrt{15}}} = 0{\cdot}162.$$

There is evidently some general relationship between the two attributes, though not a very strong one. The reader may verify for himself by using the χ^2 test that the values of C and T are significant, i.e. could not have arisen by sampling from independent attributes.

13.14. The sampling variance of the coefficient of contingency is difficult to arrive at in virtue of sheer algebraical complexity; and it is not clear how far the use of a standard error is legitimate in this connection. Reference may be made for the formulae to K. Pearson (1915*a*) and Kondo (1929). For the even more complicated question of partial contingency see K. Pearson (1915*b*).

13.15. In concluding this chapter we point out that all the measures of association and contingency discussed therein in no way rely on the possibility of the measurement of attributes on a variate-scale, or even on the possibility of arranging them in order. Rearrangement of rows and columns in the two-way tables does not affect the values of the coefficients.* In the next chapter we shall consider the relationship between variates

* Except that it may change the sign of a coefficient of association. This is equivalent to a slight change of standpoint in what is regarded as a positive association—for example, positive association between fair hair and blue eyes is equivalent to negative association between fair hair and not-blue eyes.

and certain coefficients based on the assumption that the attribute classifications are made according to the divisions of a variate-scale. These coefficients (tetrachoric r, biserial η, etc.) have been used as measures of association, but they are essentially different in character from those discussed in this chapter. The reader who refers to memoirs written on this subject between 1900 and 1920 will find it useful to remember this fact.

NOTES AND REFERENCES

The fundamental memoir on association of attributes is that of Yule (1900), who introduced the coefficient Q in it. In a later paper (1912) Yule reviewed the whole subject and proposed the coefficient denoted in this chapter by Y. This memoir contained some criticisms of Karl Pearson's coefficient now known as tetrachoric r (cf. Chapter 14) and evoked a reply from Pearson and Heron (1913) which is remarkable for having missed the point over more pages (173) than perhaps any other memoir in statistical history.

Pearson's coefficient of contingency C was introduced in 1904. Corrections to this coefficient were subsequently proposed, being based on the notion of an underlying variate (K. Pearson, 1913). For references to the other coefficients proposed on this basis, see Chapter 14.

Kondo, T. (1929), " On the standard error of the mean square contingency," *Biometrika*, **21**, 376.

Pearson, K. (1904), "On the theory of contingency and its relation to association and normal correlation," *Drapers' Company Research Memoirs, Biometric Series I*, Dulau and Co., London.

—— (1913), " On the measurement of the influence of broad categories on correlation," *Biometrika*, **9**, 116.

—— and Heron, D. (1913), " On theories of association," *Biometrika*, **9**, 159.

—— (1915*a*), " On the probable error of a coefficient of mean square contingency," *Biometrika*, **10**, 590.

—— (1915*b*), " On the general theory of multiple contingency, with special reference to partial contingency," *Biometrika*, **11**, 145.

Yule, G. U. (1900), " On the association of attributes in statistics," *Phil. Trans.*, A, **194**, 257.

—— (1912), " On the methods of measuring the association between two attributes," *Jour. Roy. Statist. Soc.*, **75**, 579.

EXERCISES

13.1. Show that the coefficient of association is greater in absolute value than the coefficient of colligation, except when both are zero or unity.

13.2. Show that for a contingency table with a constant number of rows and columns the Pearson coefficient of contingency C is equal to the Tschuprow coefficient T for two values of $\frac{\chi^2}{N}$, one of which is zero; that for $\frac{\chi^2}{N}$ between these values $C > T$, and for $\frac{\chi^2}{N}$ greater than the higher value $T > C$.

13.3. The following table shows 68 lobelia plants classified according to whether they were cross- or self-fertilised and above or below average height.

	Above Average.	Below Average.	Totals.
Cross-fertilised		17	34
Self-fertilised .	12	22	34
Totals	29	39	

Show that $Y = 0{\cdot}150$ and that this is not significant of association if these data are a random sample from lobelia plants generally.

13.4. In the hair- and eye-colour of Table **12.4** show that $C = 0{\cdot}37$ and $T = 0{\cdot}25$.

13.5. In a paper discussing whether laterality of hand is associated with laterality of eye (measured by astigmatism, acuity of vision, etc.) T. L. Woo obtained the following results (*Biometrika*, **20**A, pp. 79–148) :—

Ocular Laterality for General Astigmatism.

Manual Laterality as Determined by a Balancing Test.	"Left-eyed."			
Left-handed . .	34	62		124
Ambidextrous .	27		20	
Right-handed . .	57	105		214
Totals	118	195	100	413

Show that laterality of eye is only slightly associated with laterality of hand, and that the association is not significant.

CHAPTER 14

PRODUCT-MOMENT CORRELATION

14.1. At the end of Chapter 1 there were given a few examples of bivariate frequency tables. We now proceed to consider such tables in greater detail and to discuss methods of measuring the dependence of the two variates represented in them. It is, of course, possible to treat the problem by the methods of the previous chapter and regard the tables as contingency tables; but when data are classified according to a numerical variable more exact methods are available in an important class of cases.

The types of bivariate distribution arising in practice are not so easy to classify as the univariate types. Table 1.15 on page 20, showing the distribution of beans according to length and breadth, and Table 1.25 on page 27 showing the number of cows according to age and milk yield, evidently correspond more or less to the unimodal univariate distribution, for not only the border frequencies but the frequencies in individual rows and columns are of the unimodal type. Biometric distributions are often of this character. On the other hand, Table 1.26 on page 28, showing discount rates and bank reserves, has the border column of the unimodal type and the border row of the J-shaped type. In Tables 14.1 to 14.3 are given three more examples of the kind of material encountered in practice. Table 14.1 shows the distribution of a number of persons according to age and highest audible pitch; Table 14.2 the distribution of registration districts according to proportion of male births and total number of births; and Table 14.3 shows the distribution of sons according to stature of son and stature of father.

TABLE

stribution of 3379 P ccording to Age and Highest Audible P

(From Y. Koga and . M. Morant, *Biometrika*, 15, 346.)

The numbers in brackets are explained in Example 14.1, p. 331

Ag

Highest audible pitch, thousand vibrations per second.	5·5–	8·5–	11·5–	14·5–	17·5–	20·5–	23·5–	26·5–	29·5–	32·5–	35·5–	38·5–	41·5–	44·5–	47·5–	50·5–	53·5–	56·5–	59·5–	62·5–	65·5–	68·5–	71·5–	74·5–	77·5	TOTALS
5–	—	—	—	—	—	1 (0)	—	1 (−14)	—	—	—	—	—	—	—	—	—	1 (−84)	—	—	—	—	—	—	—	3
7–	—	—	1 (18)	1 (12)	3 (6)	2 (0)	—	3 (−12)	1 (−18)	1 (−24)	—	2 (−36)	1 (−42)	3 (−48)	1 (−54)	3 (−60)	3 (−66)	5 (−72)	2 (−78)	4 (−84)	3 (−90)	2 (−96)	1 (−102)	3 (−108)	—	45
9–	—	—	—	—	—	—	1 (−5)	—	—	—	1 (−25)	—	—	2 (−40)	2 (−45)	—	1 (−55)	1 (−60)	—	—	1 (−75)	—	—	—	1 (−95)	10
11–	—	—	1 (12)	2 (8)	2 (4)	3 (0)	2 (−4)	4 (−8)	2 (−12)	4 (−16)	2 (−20)	5 (−24)	7 (−28)	6 (−32)	12 (−36)	14 (−40)	5 (−44)	11 (−48)	6 (−52)	8 (−56)	3 (−60)	2 (−64)	—	2 (−72)	1 (−76)	104
13–	—	—	2 (9)	3 (6)	7 (3)	2 (0)	1 (−3)	4 (−6)	7 (−9)	5 (−12)	10 (−15)	8 (−18)	4 (−21)	6 (−24)	11 (−27)	4 (−30)	6 (−33)	2 (−36)	4 (−39)	4 (−42)	2 (−45)	—	1 (−51)	—	—	93
15–	—	2 (8)	7 (6)	21 (4)	40 (2)	32 (0)	33 (−2)	27 (−4)	18 (−6)	20 (−8)	21 (−10)	22 (−12)	19 (−14)	15 (−16)	16 (−18)	6 (−20)	5 (−22)	3 (−24)	1 (−26)	1 (−28)	1 (−30)	—	—	—		310
17–	3 (5)	4 (4)	26 (3)	49 (2)	121 (1)	105 (0)	72 (−1)	50 (−2)	36 (−3)	28 (−4)	21 (−5)	21 (−6)	18 (−7)	13 (−8)	4 (−9)	3 (−10)	2 (−11)	—	—	—	—	—	—	—	—	576
19	3 (0)	23 (0)	74 (0)	153 (0)	261 (0)	178 (0)	110 (0)	103 (0)	47 (0)	34 (0)	25 (0)	18 (0)	13 (0)	3 (0)	3 (0)	1 (0)	—	1 (0)	1 (0)	—	—	—	—	—	—	1051
21	9 (−5)	39 (−4)	112 (−3)	158 (−2)	303 (−1)	164 (0)	68 (1)	47 (2)	24 (3)	21 (4)	6 (5)	2 (6)	3 (7)	1 (8)	—	—	—	—		—	—	—	—	—	—	957
23	—	10 (−8)	27 (−6)	37 (−4)	50 (−2)	22 (0)	11 (2)	4 (4)	2 (6)	1 (8)	1 (10)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	165
25	—	4 (−12)	5 (−9)	14 (−6)	11 (−3)	3 (0)	2 (3)	1 (6)	—	1 (12)	—	—	—	—	—	—	—	—		—	—	—	—	—	—	41
27		1 (−16)	4 (−12)	3 (−8)	6 (−4)	1 (0)	—	—	1 (12)	—	—	—	—	—	—	—	—	—		—	—	—	—	—	—	16
29				1 (−10)	—	1 (0)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
31	—		2 (−18)				—			—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
33		1 (−28)	1 (−21)		1 (−7)		1 (7)	—		—	—	—	—	—	—	—	—		—	—	—	—	—	—	—	4
TOTALS	15	84	262	442	805	514	301	244	138	115	87	78	65	49	49	31	22	24	14	17	10	4	2	5	2	3379

TABLE 14.2

Showing the Number of Registration Districts in England and Wales exhibiting (1) *a given Proportion of Male Births*, (2) *a given Total Number of Births during the Decade 1881–90.* (The Data as to Total Births and Numbers of Male and Female Births from *Decennial Supplement to Report of the Registrar-General.* Table from H. D. Vigor and G. U. Yule, *Jour. Roy. Stat. Soc.*, **69**, 1906.)

(2) Total Number of Births in District (000's omitted) during Decade.	(1) Proportion of Male Births per 1000 of all Births.																											
	465–67.	468–70.	471–73.	474–76.	477–79.	480–82.	483–85.	486–88.	489–91.	492–94.	495–97.	498–500.	501–03.	504–06.	507–09.	510–12.	513–15.	516–18.	519–21.	522–24.	525–27.	528–30.	531–33.	534–36.	537–39.	540–42.	543–45.	TOTALS.
0– 4	1	—	1	—	—	2	1	2	2	4	8	9	18	19	12	21	14	12	9	5	3	1	1	2	—	1	1	149
4– 8	—	—	—	—	—	—	1	—	2	2	5	7	20	27	29	42	36	18	10	4	—	—	—	1	—	—	—	204
8– 12	—	—	—	—	—	—	—	—	—	—	2	1	5	15	17	18	16	7	4	1	—	—	—	—	—	—	—	86
12– 16	—	—	—	—	—	—	—	—	—	—	—	—	1	10	8	12	8	6	1	1	1	—	—	—	—	—	—	48
16– 20	—	—	—	—	—	—	—	—	—	—	—	—	—	5	6	9	4	1	1	—	—	—	—	—	—	—	—	26
20– 24	—	—	—	—	—	—	—	—	—	—	—	—	—	1	5	4	3	1	1	—	—	—	—	—	—	—	—	15
24– 28	—	—	—	—	—	—	—	—	—	—	—	—	2	3	5	3	1	1	—	—	—	—	—	—	—	—	—	15
28– 32	—	—	—	—	—	—	—	—	—	—	—	1	—	2	6	2	1	—	—	—	—	—	—	—	—	—	—	12
32– 36	—	—	—	—	—	—	—	—	—	—	—	—	—	3	3	—	1	—	—	—	—	—	—	—	—	—	—	7
36– 40	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4	—	2	—	—	—	—	—	—	—	—	—	—	6
40– 44	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	4	1	—	—	—	—	—	—	—	—	—	—	8
44– 48	—	—	—	—	—	—	—	—	—	—	—	—	—	2	5	2	1	1	—	—	—	—	—	—	—	—	—	11
48– 52	—	—	—	—	—	—	—	—	—	—	—	—	—	2	6	3	—	—	—	—	—	—	—	—	—	—	—	11
52– 56	—	—	—	—	—	—	—	—	—	—	—	—	—	2	5	—	—	—	—	—	—	—	—	—	—	—	—	7
56– 60	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	3
60– 64	—	—	—	—	—	—	—	—	—	—	—	—	—	3	1	1	—	—	—	—	—	—	—	—	—	—	—	5
64– 68	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	2	—	—	—	—	—	—	—	—	—	—	—	3
68– 72	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	1	—	—	—	—	—	—	—	—	—	—	—	3
72– 76	—	—	—	—	—	—	—	—	—	—	—	—	—	1	3	—	—	—	—	—	—	—	—	—	—	—	—	4
76– 80	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
80– 84	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
84– 88	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	2
88– 92	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	2
92– 96	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
96–100	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	2
100– 04	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
104– 08	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
148– 52	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
TOTALS	1	—	1	—	—	2	2	2	4	6	15	18	46	98	125	129	88	47	26	11	4	1	1	3	—	1	1	632

14.2. In accordance with the definitions of the previous chapter we may say that two variables are independent in a bivariate table if the observed frequency in the jth row and ith column (A_iB_j), is equal to $\frac{(A_i)(B_j)}{N}$. In such a case, for any two rows j and k the frequencies will be proportional to $(A_i)(B_j)$ and $(A_i)(B_k)$; so that the distribution in any row is similar and similarly situated to that in any other row. It will, for example, have the same mean and variance. Similarly for the columns.

The measures of dependence we shall consider are related to the extent to which row and column means and variances differ among themselves. If the variates are independent all row means and all column means are equal. The converse is not true in general, but becomes so in an important special case when the distribution is normal.

TABLE 14.3

Distribution of 1078 Sons according to (1) Stature of Father and (2) Stature of Son: One or Two Sons only of each Father.

(From Karl Pearson and Alice Lee, *Biometrika*, **2** (1903), 415.)

Measurements in inches. Note that where a height falls on the border-line of an interval, one-half of the individual is assigned to each contiguous interval.

(2)	(1) Stature of Father.																	Total
	58·5–59·5	59·5–60·5	60·5–61·5	61·5–62·5	62·5–63·5	63·5–64·5	64·5–65·5	65·5–66·5	66·5–67·5	67·5–68·5	68·5–69·5	69·5–70·5	70·5–71·5	71·5–72·5	72·5–73·5	73·5–74·5	74·5–75·5	
59·5–60·5	—	—	—	—	0·5	0·5	1	—	—	—	—	—	—	—	—	—	—	2
60·5–61·5	—	—	—	—	0·5	—	—	—	1	—	—	—	—	—	—	—	—	1·5
61·5–62·5	—	0·25	0·25	—	0·5	1	0·25	0·25	0·5	0·5	—	—	—	—	—	—	—	3·5
62·5–63·5	—	0·25	0·25	2·25	2·25	2	4	5	2·75	1·25	—	0·25	0·25	—	—	—	—	20·5
63·5–64·5	1	—	1·5	3·75	3	4·25	8	9·25	3	1·25	1·5	0·75	1·25	—	—	—	—	38·5
64·5–65·5	2	1	0·5	2	3·25	9·5	13·5	10·75	7·5	5·5	3·5	2·5	—	—	—	—	—	61·5
65·5–66·5	—	0·5	1	2·25	5·25	9·5	10	16·75	17·5	16	5·25	2	2·5	1	—	—	—	89·5
66·5–67·5	—	1·5	2	4·75	3·5	13·75	19·75	26·5	25·75	19·5	12·5	13·75	5·25	0·5	1	—	—	148
67·5–68·5	—	—	1·5	2	7·5	10	10·25	24·25	31·5	23·5	29·5	13·25	8·5	9·5	2·25	—	—	178·5
68·5–69·5	—	—	1	—	5·25	5	12·75	18·25	16	24	29	21·5	10	3·5	2·25	—	1	149·5
69·5–70·5	—	—	—	—	1	2·5	5·75	18·75	11·75	19·5	22·5	19·5	14·5	6·25	3·5	1·5	1	128
70·5–71·5	—	—	—	—	—	3·25	5	8·75	10·75	19	14·75	20·75	10·75	8	5	1	1	108
71·5–72·5	—	—	—	—	—	0·25	3	1·25	7	7·75	10·75	11·25	10	8·5	2·75	0·5	—	63
72·5–73·5	—	—	—	—	—	—	0·75	0·75	2·5	7·5	6·5	6	7·5	6·25	3·25	0·5	0·5	42
73·5–74·5	—	—	—	—	1	—	1·5	1·5	—	5·25	2·25	2·5	6·5	3·25	3·25	—	2	29
74·5–75·5	—	—	—	—	—	—	—	—	—	1	2	—	2·5	0·75	1·75	0·5	—	8·5
75·5–76·5	—	—	—	—	—	—	—	—	—	1·25	0·25	—	0·5	1	1	—	—	4
76·5–77·5	—	—	—	—	—	—	—	—	—	1·25	0·25	1	—	—	1·5	—	—	4
77·5–78·5	—	—	—	—	—	—	—	—	—	—	1	1	—	0·25	0·75	—	—	3
78·5–79·5	—	—	—	—	—	—	—	—	—	—	—	—	—	0·25	0·25	—	—	0·5
Totals	3	3·5	8	17	33·5	61·5	95·5	142	137·5	154	141·5	116	78	49	23·5	4	5·5	1078

Regression

14.3. The rows and columns will be referred to by the general term "arrays" and we shall consider the two variates x and y, x being taken to vary horizontally, i.e. in rows, and y vertically, i.e. in columns. Consider then the means of arrays. Take two axes OX and OY at right angles representing the variates x and y. On this frame of reference plot the points whose abscissae are the centres of the x-intervals and whose ordinates are the means of the corresponding distributions of y in the columns centred at the appropriate x's. Similarly, plot the means of the x-distributions against the centres of the corresponding y-intervals. (In practice it is useful to distinguish the two, the means of x's being denoted by small circles and those of y's by crosses.) Fig. 14.1 shows such a diagram for the data of Table 14.3 and Fig. 14.2 for the data of Table 14.2.

The means of arrays will in general lie more or less closely round smooth curves. For example, in Fig. 14.1 they lie approximately on straight lines, whereas in Fig. 14.2 one set of means certainly does not. Such curves are called regression curves and their equations with respect to OX and OY are called regression equations. If the lines are straight the regression is said to be linear; if not, curvilinear or skew.

To put these geometrically expressed ideas in analytical language, suppose the mean

of x in the array centred at y_i is $\bar{x}_i$. Then the points $(\bar{x}_i, y_i)$ may be represented by a functional equation—

$$\bar{x} = f(y) \qquad \ldots \qquad (14.1)$$

which is the regression equation of x on y. If the regression is linear,

$$\bar{x} = \beta y + \alpha.$$

There will also be an equation

$$\bar{y} = g(x), \qquad \ldots \qquad (14.2)$$

the regression of y on x.

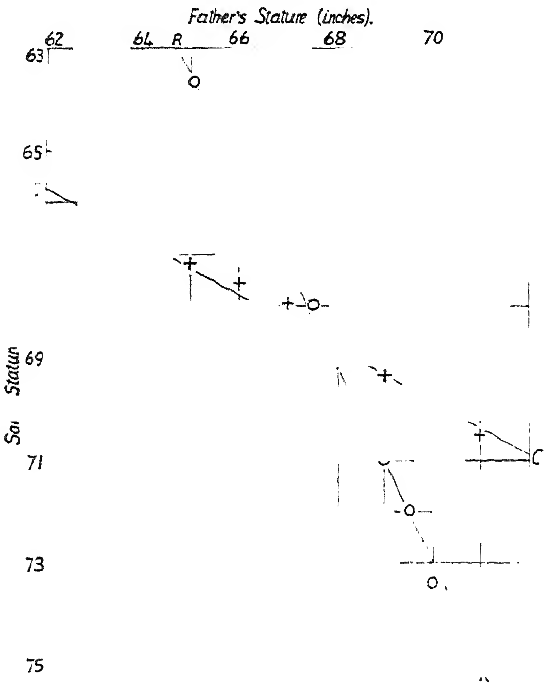


FIG. 14.1.—Regression Lines of Data of Table 14.3. Means of rows shown by circles, and corresponding regression line by RR; means of columns shown by crosses, and corresponding regression line by CC.

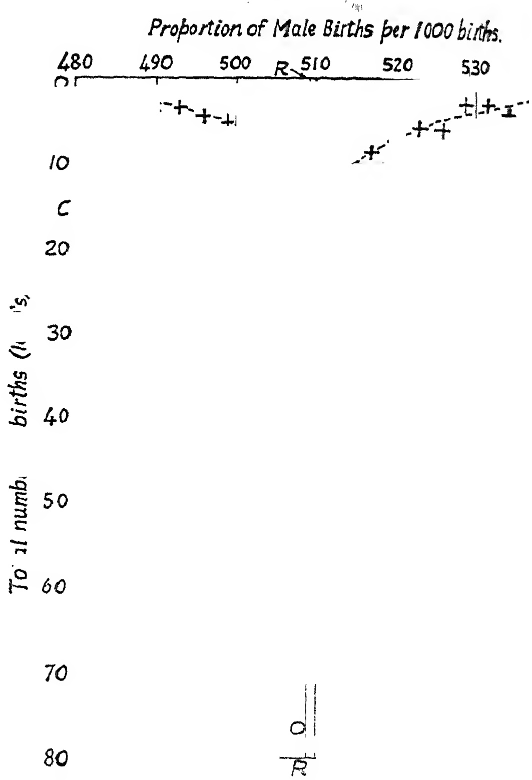


FIG. 14.2.—Regression Lines of Data of Table 14.2. Means of rows shown by circles, and corresponding regression line by RR; means of columns shown by crosses, and regression line by CC.

In this chapter we shall mainly be concerned with the case in which regressions are linear or very nearly so.

14.4. In an observed distribution the means of arrays will not as a rule lie exactly on straight lines, or indeed on any simple curves, although they may be very near to doing so. The question then arises: if the regression is "approximately" linear, what is the best line to take as the regression line? The question may be answered by an appeal to the method of least squares. The regression of x on y, say $\bar{x} = \beta y + \alpha$, will be determined by minimising the sum

$$U = \Sigma N_i(\bar{x}_i - \beta y_i - \alpha)^2, \qquad (14.3)$$

the summation extending over all y-intervals. Here N_i represents the frequency in the ith row, and we note that $\Sigma(N_i \bar{x}_i)$ is equal to the total frequency N times the mean of x for the whole distribution.

From (14.3) we have, for the minimal values of α and β,

$$\frac{\partial U}{\partial \alpha} = -2\Sigma N_i(\bar{x}_i - \beta y_i - \alpha) = 0 \qquad (14.4)$$

$$\frac{\partial U}{\partial \beta} = 2\Sigma N_i y_i(\bar{x}_i - \beta y_i - \alpha) = 0 \qquad (14.5)$$

Now choose the origin at the means of x and y for the distribution. Then $\Sigma N_i \bar{x}_i = 0$ and $\Sigma N_i y_i = 0$ and hence, from (14.4)

$$\alpha = 0.$$

From (14.5) we then have

$$\Sigma(N_i y_i \bar{x}_i) - \beta \Sigma(N_i y_i^2) = 0.$$

Now since the origin has been chosen at the mean of x and y, $\Sigma(N_i y_i \bar{x}_i) = N \operatorname{cov}(x, y)$ and $\Sigma(N_i y_i^2) = N \operatorname{var} y$. Hence we have

$$\beta = \frac{\operatorname{cov}(x, y)}{\operatorname{var} y}. \qquad (14.6)$$

The equation of the regression of x on y, taking X and Y to be current co-ordinates, will then be

$$X = \frac{\operatorname{cov}(x, y)}{\operatorname{var} y} Y. \qquad (14.7)$$

Referred to an arbitrary origin for which the means of x, y are $\bar{x}$, $\bar{y}$, the equation is

$$(X - \bar{x}) = \frac{\operatorname{cov}(x, y)}{\operatorname{var} y}(Y - \bar{y}). \qquad (14.8)$$

Similarly we find for the regression of y on x

$$(Y - \bar{y}) = \frac{\operatorname{cov}(x, y)}{\operatorname{var} x}(X - \bar{x}). \qquad (14.9)$$

Equations (14.8) and (14.9) are fundamental. If the regressions are exactly linear they give the regression equations; if not, they give the "best" straight regression lines in the sense of the method of least squares.

The Coefficient of Product-moment Correlation

14.5. The coefficients $\dfrac{\operatorname{cov}(x, y)}{\operatorname{var} y}$ and $\dfrac{\operatorname{cov}(x, y)}{\operatorname{var} x}$ are called regression coefficients and will be denoted by β_1 and β_2 respectively.*

We now define

$$\rho = (\beta_1\beta_2)^{\frac{1}{2}}$$
$$= \frac{\operatorname{cov}(x, y)}{(\operatorname{var} x \operatorname{var} y)^{\frac{1}{2}}} \qquad (14.10)$$

* There is little danger of confusion between this notation and the use of β_1, β_2 to indicate measures of skewness and kurtosis. The two rarely occur in the same context.

ρ is called the coefficient of product-moment correlation or briefly the correlation coefficient. It provides an important measure of the relation between two variates for which the regressions are approximately linear. In this expression the square root is to have positive sign.

Let us note in the first place that ρ cannot be greater than unity in absolute value. For we have, taking an origin at the means and summing for all pairs of values of x, y over the population,

$$\begin{aligned}\Sigma(x - \beta_1 y)^2 &= \Sigma(x^2) - 2\beta_1\Sigma(xy) + \beta_1^2\Sigma(y^2) \\ &= \Sigma(x^2)\left\{1 - \frac{2\beta_1\Sigma(xy)}{\Sigma(x^2)} + \frac{\beta_1^2\Sigma(y^2)}{\Sigma(x^2)}\right\} \\ &= \Sigma(x^2)\{1 - 2\beta_1\beta_2 + \beta_1\beta_2\} \\ &= \Sigma(x^2)(1 - \rho^2) \quad . \quad . \qquad (14.11)\end{aligned}$$

Thus $1 - \rho^2$ cannot be negative.

Furthermore, if $\rho = \pm 1$, $\Sigma(x - \beta_1 y)^2 = 0$ and hence every $x - \beta_1 y = 0$. Thus the variates are linearly related by the equation $X - \beta_1 Y = 0$. If $\rho = 0$ the regression equations become $X = 0$, $Y = 0$, for then cov $(x, y) = 0$. Hence the means of arrays are the same for all arrays.

14.6. If $\rho = +1$ we say that the variates are perfectly positively correlated; if $0 < \rho < 1$, that they are positively correlated; if $\rho = 0$, that they are uncorrelated; if $0 > \rho > -1$, that they are negatively correlated; and if $\rho = -1$, that they are perfectly negatively correlated.

"Uncorrelated" is not the same thing as "independent." If the variates are independent they are uncorrelated, but not *vice-versa*. Table 14.2 and Fig. 14.2 illustrate this point. The regression lines, as shown in the figure, are close to $X = 0$, $Y = 0$ and the correlation is, in fact, very small ($-$ 0·014). But the variates are obviously far from independence and if the data are grouped, in columns up to 494·5, by single columns up to 521·5, and over 521·5, and by rows 8–11, 12–13, 14–15, 16–17, 18–19, 20–21, 22–23, 24–25, 26–27, 28 and over, the coefficient of contingency is 0·47.

14.7. The calculation of the correlation coefficient in numerical examples requires that of the means and variances of the two variates and their covariance. The last is the only new type appearing, the others being calculable from border frequencies in the manner exemplified in Chapter 3.

Taking an arbitrary origin, we have, if the means of x and y are a and b.

$$\begin{aligned}N \text{ cov } (x, y) &= \Sigma(x - a)(y - b) \\ &= \Sigma(xy) - a\Sigma(y) - b\Sigma(x) + Nab \\ &= \Sigma(xy) - Nab\end{aligned}$$

$$\text{cov } (x, y) = \frac{1}{N}\Sigma(xy) - ab. \qquad (14.12)$$

Thus we may, as in the univariate case, take an arbitrary origin for arithmetical convenience, calculate the product sum $\Sigma(xy)$ and determine the covariance by the use of (14.12). The

calculation of the product sum is exemplified below. As seen in **3.30**, no Sheppard's corrections are required for the first product moment.

Example 14.1

To find the correlation coefficient and regression lines for the data of Table 14.1.

We find the means and variances from the border columns in the usual way. An arbitrary mean is taken for x (age) at the centre of the interval 20·5 years and for y (highest pitch) at the centre of the interval 19– thousand vibrations which may be taken as 19,995. We find

$$\begin{aligned} \Sigma(x) &= 2,604, & \mu_1'(x) &= 0{\cdot}770,642 \\ \Sigma(y) &= -708, & \mu_1'(y) &= -0{\cdot}209,529 \\ \Sigma(x^2) &= 47,392, & \mu_2(x) &= 13{\cdot}348,229 \\ \Sigma(y^2) &= 8894, & \mu_2(y) &= 2{\cdot}504,904. \end{aligned}$$

To find the product sum $\Sigma(xy)$ we require the product xy for each non-zero cell of the table. These products are shown in brackets in Table 14.1. Then we have, reading the table from left to right and from top to bottom,

$$\begin{aligned} \Sigma(xy) &= (1 \times 0) + (1 \times -14) + (1 \times -84) \\ &\quad + (1 \times 18) + (1 \times 12) + (3 \times 6) + \text{etc.} \\ &= -12,535, \end{aligned}$$

whence

$$\begin{aligned} \text{cov}\,(x, y) &= \frac{-12,535}{3379} - (0{\cdot}770,642)(-0{\cdot}209,529) \\ &= -3{\cdot}548,205. \end{aligned}$$

Thus

$$\rho = \frac{\text{cov}\,(x, y)}{\{\mu_2(x)\,\mu_2(y)\}^{\frac{1}{2}}} = -0{\cdot}6136,$$

a substantial negative correlation. The highest audible pitch decreases with increasing age.

We also find

$$\beta_1 = \frac{\text{cov}\,(x, y)}{\text{var}\,y} = -1{\cdot}417$$

$$\beta_2 = \frac{\text{cov}\,(x, y)}{\text{var}\,x} = -0{\cdot}2658.$$

The regression equations, for the units of the table and with our arbitrary means, are then

$$\begin{aligned} X - 0{\cdot}7706 &= -1{\cdot}417(Y + 0{\cdot}2095) \\ Y + 0{\cdot}2095 &= -0{\cdot}2658(X - 0{\cdot}7706). \end{aligned}$$

Example 14.2

The following device is often useful in calculating product moments. We recall that

$$\begin{aligned} 2\Sigma(xy) &= \Sigma(x + y)^2 - \Sigma(x^2) - \Sigma(y^2) \\ &= \Sigma(x^2) + \Sigma(y^2) - \Sigma(x - y)^2. \end{aligned}$$

Thus we may find $\Sigma(xy)$ from either $\Sigma(x + y)^2$ or $\Sigma(x - y)^2$, and these quantities are often more convenient to calculate.

For example, in the preceding example we note that $x + y$ is constant down the diagonals running from the bottom right-hand to the top left-hand corner of the table. Taking $x + y$ to be zero in the cell centred at $x = 20{\cdot}5$– and $y = 19$–, we may, in our

units, take it to be + 1 in the cell 23·5–, 19–, and –1 in 17·5–, 19–, and so on. If we sum up the diagonals we get—

$x + y$.	Sum.	$x + y$.	Sum.
–9	1	4	124
–8	1	5	112
–7	5	6	90
–6	11	7	59
–5	20	8	38
–4	93	9	23
–3	207	10	21
–2	434	11	9
–1	594	12	9
0	637	13	2
1	418	14	4
2	281	15	1
3	185		
		TOTAL	3379

The total is 3379, which provides a check on the work. We then find the sum of squares in the ordinary way, obtaining

$$\Sigma(x + y)^2 = 31{,}216 = \Sigma(x^2) + \Sigma(y^2) + 2\,\Sigma(xy).$$

Then
$$\Sigma(xy) = \tfrac{1}{2}(31{,}216 - 47{,}392 - 8894)$$
$$= -\ 12{,}535 \text{ as before.}$$

The rest of the calculation follows the same lines as in Example 14.1.

We should have obtained the same result for $\Sigma(xy)$ if we had summed up the other diagonal. Which diagonal is chosen depends on how the frequencies lie in the table.

Example 14.3

In the foregoing the regression lines and the correlation coefficient were arrived at from a consideration of grouped frequencies in a bivariate table. We may, however, apply the same ideas to ungrouped material. There are no longer means of arrays, but the regression lines are still to be interpreted as the lines of best fit to the N pairs of variate-values and the correlation coefficient as a measure of relationship between variates.

Table 14.4 shows the yields of wheat and potatoes in 48 counties of England in 1936. In this particular case it is hardly worth while taking an arbitrary origin other than that given. We find (x = wheat, y = potatoes)

$$\begin{aligned}
\Sigma(x) &= 758{\cdot}0, & \mu_1'(x) &= 15{\cdot}791{,}667\\
\Sigma(y) &= 291{\cdot}1, & \mu_1'(y) &= 6{\cdot}064{,}583\\
\Sigma(x^2) &= 12{,}170{\cdot}48, & \mu_2(x) &= 4{\cdot}174{,}930\\
\Sigma(y^2) &= 1791{\cdot}03, & \mu_2(y) &= 0{\cdot}533{,}958\\
\Sigma(xy) &= 4612{\cdot}64, & \mu_{11}(x, y) &= 0{\cdot}326{,}888
\end{aligned}$$

$$r = \frac{0{\cdot}326{,}888}{\sqrt{(4{\cdot}174{,}930 \times 0{\cdot}533{,}958)}}$$
$$= 0{\cdot}2189.$$
$$\beta_1 = 0{\cdot}6122,$$
$$\beta_2 = 0{\cdot}07830.$$

TABLE 14.4

Yields of Wheat and Potatoes in 48 Counties in England in 1936.

County.	Wheat (cwts. per acre).	Potatoes (tons per acre).	County.	Wheat (cwts. per acre).	Potatoes (tons per acre).
Bedfordshire	16·0	5·3	Northamptonshire	14·3	4·9
Huntingdonshire	16·0	6·6	Peterborough	14·4	5·6
Cambridgeshire	16·4	6·1	Buckinghamshire	15·2	6·4
Ely	20·5	5·5	Oxfordshire	14·1	6·9
Suffolk, West	18·2	6·9	Warwickshire	15·4	5·6
Suffolk, East	16·3	6·1	Shropshire	16·5	6·1
Essex	17·7	6·4	Worcestershire	14·2	5·7
Hertfordshire	15·3	6·3	Gloucestershire	13·2	5·0
Middlesex	16·5	7·8	Wiltshire	13·8	6·5
Norfolk	16·9	8·3	Herefordshire	14·4	6·2
Lincs. (Holland)	21·8	5·7	Somersetshire	13·4	5·2
,, (Kesteven)	15·5	6·2	Dorsetshire	11·2	6·6
,, (Lindsey)	15·8	6·0	Devonshire	14·4	5·8
Yorkshire (East Riding)	16·1	6·1	Cornwall	15·4	6·3
Kent	18·5	6·6	Northumberland	18·5	6·3
Surrey	12·7	4·8	Durham	16·4	5·8
Sussex, East	15·7	4·9	Yorkshire (North Riding)	17·0	5·9
Sussex, West	14·3	5·1	,, (West Riding)	16·9	6·5
Berkshire	13·8	5·5	Cumberland	17·5	5·8
Hampshire	12·8	6·7	Westmorland	15·8	5·7
Isle of Wight	12·0	6·5	Lancashire	19·2	7·2
Nottinghamshire	15·6	5·2	Cheshire	17·7	6·5
Leicestershire	15·8	5·2	Derbyshire	15·2	5·4
Rutland	16·6	7·1	Staffordshire	17·1	6·3

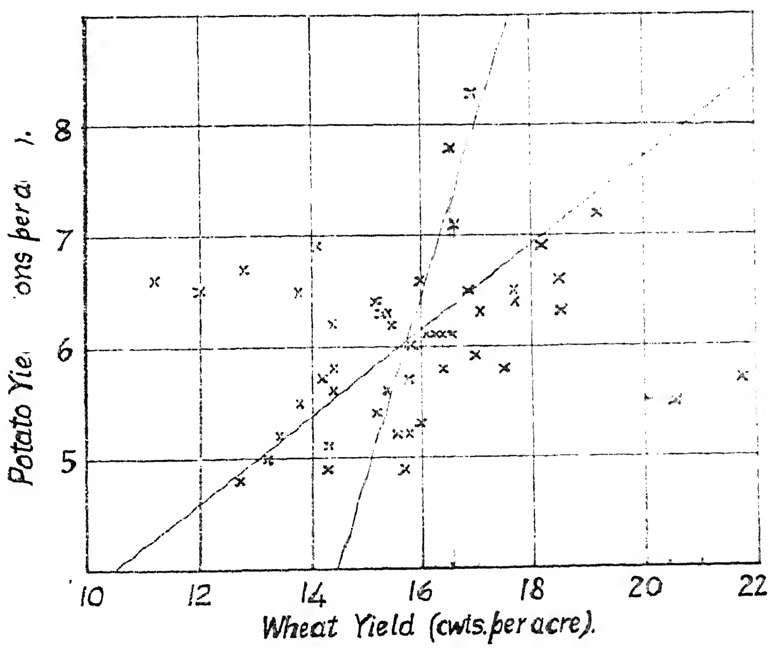


FIG. 14.3.—Scatter Diagram of the Data of Table 14.4.

The regression lines are

$$X - 15{\cdot}792 = 0{\cdot}6122\ (Y - \ \ 6{\cdot}065)$$
$$Y - \ \ 6{\cdot}065 = 0{\cdot}0783\ (X - 15{\cdot}792)$$

The data are shown in a graphical form in Fig. 14.3. Corresponding to each pair of values (x, y) there is plotted a point with those values as abscissa and ordinate. The totality of points furnishes what is known, for obvious reasons, as a scatter diagram. The two regression lines are also shown.

The Bivariate Normal Distribution

14.8. The distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right\} dx\, dy \quad . \quad (14.13)$$

has already arisen (**5.24**) as the natural extension to two variates of the univariate normal distribution. In writing ρ in (14.13) we have anticipated a result which will now be proved, namely that ρ in that equation is in fact the correlation coefficient of the distribution.

The characteristic function of (14.13) is

$$\phi(t, u) = \exp[-\tfrac{1}{2}(t^2\sigma_1^2 + 2ut\rho\sigma_1\sigma_2 + u^2\sigma_2^2]$$

whence

$$\operatorname{var} x = \sigma_1^2, \qquad \operatorname{var} y = \sigma_2^2$$
$$\operatorname{cov}(x, y) = \rho\sigma_1\sigma_2$$

and the correlation coefficient is $\dfrac{\rho\sigma_1\sigma_2}{\sqrt{(\sigma_1^2\sigma_2^2)}} = \rho$, as stated.

The exponent in (14.13) may be written

$$-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_1} - \frac{\rho y}{\sigma_2}\right)^2 + \frac{y^2}{\sigma_2^2}(1-\rho^2)\right\} \quad . \quad (14.14)$$

$$= -\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y}{\sigma_2} - \frac{\rho x}{\sigma_1}\right)^2 + \frac{x^2}{\sigma_1^2}(1-\rho^2)\right\} \quad . \quad (14.15)$$

Thus for any fixed y, x is distributed normally about a mean given by

$$\frac{x}{\sigma_1} = \frac{\rho y}{\sigma_2},$$

and hence the means of the x-arrays of infinite thinness lie on the line

$$\frac{X}{\sigma_1} = \frac{\rho Y}{\sigma_2} \quad . \quad . \quad . \quad . \quad (14.16)$$

and this is the regression of x on y. Similarly the means of y-arrays lie on

$$\frac{Y}{\sigma_2} = \frac{\rho X}{\sigma_1}, \quad . \quad . \quad . \quad . \quad (14.17)$$

the regression of y on x. Thus the regression lines of the bivariate normal surface are exactly linear.

Furthermore, from (14.14) it is seen that the variance of an array of x for fixed y is

$$\sigma_1^2(1-\rho^2),$$

i.e. is independent of y. Similarly the variance of an array of y is

$$\sigma_2^2(1-\rho^2)$$

and is independent of x.

Thus the variance of x-arrays is the same for all arrays; and so for y. Distributions for which this is true are called homoscedastic.

If $\rho = 0$ the distribution becomes the product of two normal distributions and the variates are independent. Thus two uncorrelated *normal* variates are independent.

14.9. A criterion for linearity of regression for the general bivariate surface may be obtained in terms of moments or cumulants. Taking an origin at the means of the two variates we have, if the regression of x on y is linear,

$$\bar{x} = \beta_1 y$$

or, if $f(x, y)$ is the frequency function,

$$\int_{-\infty}^{\infty} x\, f(x, y)\, dx = \beta_1 y \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Multiplying both sides by y^p and integrating over the range of y, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^p\, x\, f(x, y)\, dx\, dy = \beta_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^{p+1} f(x, y)\, dx\, dy$$

or $$\mu_{1,p} = \beta_1 \mu_{0,p+1}. \qquad (14.18)$$

In particular $$\mu_{11} = \beta_1 \mu_{02},$$

so that $$\mu_{02}\mu_{1,p} = \mu_{11}\mu_{0,p+1}, \qquad (14.19)$$

a condition on the central moments. Recalling the definition of bivariate cumulants

$$\left(\kappa_{ij}\frac{t_1^i t_2^j}{i!\, j!}\right) = \log \Sigma\left(\mu_{ij}\frac{t_1^i t_2^j}{i!\, j!}\right)$$

we see that the univariate mean moments are related to cumulants in the same way as moments $\mu_{1,p}$ to cumulants $\kappa_{1,p}$ and hence we also have

$$\kappa_{1,p} = \beta_1 \kappa_{0,p+1} \qquad (14.20)$$
$$\kappa_{02}\kappa_{1,p} = \kappa_{11}\kappa_{0,p+1} \qquad (14.21)$$

Similarly, if the regression of y on x is linear we shall have

$$\left.\begin{aligned} \mu_{20}\mu_{p,1} &= \mu_{11}\mu_{p+1,0} \\ \kappa_{20}\kappa_{p,1} &= \kappa_{11}\kappa_{p+1,0} \end{aligned}\right\} \qquad (14.22)$$

Under certain conditions these equations are *sufficient* for linearity of regression. For instance, if (14.18) is true for all p, then

$$y^p\, dy \qquad (x - \beta_1 y)\, f(x, y)\, dx \} = 0.$$

The expression in curly brackets is thus a function, not necessarily positive, whose moments vanish, and under certain general conditions this implies that the function itself vanishes, i.e. that

$$\int_{-\infty}^{\infty} (x - \beta_1 y)\, f(x, y)\, dx = 0$$

or $$\bar{x} = \beta_1 y,$$

so that the regression of x on y is linear.

Example 14.4 (from Wicksell, *Biometrika*, **25**, 126)

Consider the bivariate distribution of the squares of variates x, y which are distributed in the bivariate normal form. The characteristic function of these variates is proportional to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{itx^2+iuy^2} \exp -\frac{1}{2(1-\rho^2)}\left\{\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right\} dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp -\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_1^2}\{1 - 2\sigma_1^2(1-\rho^2)it\} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\{1 - 2\sigma_2^2(1-\rho^2)iu\}\right] dx\, dy.$$

This is proportional to (compare Exercise 1.5)

$$\begin{vmatrix} \frac{1}{\sigma_1^2}\{1 - 2\sigma_1^2(1-\rho^2)it\} & \frac{\rho}{\sigma_1\sigma_2} \\ \frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\{1 - 2\sigma_2^2(1-\rho^2)iu\} \end{vmatrix}^{-\frac{1}{2}}$$

which, except for constants, reduces to

$$\{(1 - 2\sigma_1^2 it)(1 - 2\sigma_2^2 iu) - 4\rho^2\sigma_1^2\sigma_2^2(it)(iu)\}^{-\frac{1}{2}}.$$

This is the characteristic function, for when $t = u = 0$ it reduces to unity.

Now the frequency-distribution represented by this function is evidently not normal; but its regressions are linear, for we have, taking logarithms,

$$\Sigma\left\{\kappa_{rs}\frac{(it)^r(iu)^s}{r!s!}\right\} = -\tfrac{1}{2}\log\{(1 - 2\sigma_1^2 it)(1 - 2\sigma_2^2 iu) - 4\rho^2\sigma_1^2\sigma_2^2(it)(iu)\}$$

giving, on identifying coefficients,

$$\begin{aligned}\kappa_{p,0} &= \tfrac{1}{2}(p-1)!(2\sigma_1^2)^p \\ \kappa_{p,1} &= 2\rho^2\sigma_1^2\sigma_2^2(2\sigma_1^2)^{p-1}p! \\ \kappa_{11} &= 2\rho^2\sigma_1^2\sigma_2^2\,;\end{aligned}$$

and hence

$$\kappa_{20}\,\kappa_{p,1} = \kappa_{11}\,\kappa_{p+1,0}.$$

Sampling of Regression and Correlation Coefficients

14.10. We now turn to consider the sampling problems associated with the coefficients of correlation and regression.

First of all, as to standard errors. In Example 9.6 on page 211 we have anticipated the determination of the sampling variance of the correlation coefficient itself, obtaining the result for the normal case

$$\text{var}\, r = \frac{1}{n}(1-\rho^2)^2 \qquad . \quad (14.23)$$

Here, as usual, the Roman r is written for the value of ρ in the sample and n is the number in the sample. The result of (14.23) is not of great value, since the distribution of r tends to normality very slowly if ρ is not close to zero. It is probably as well not to use (14.23) unless n is greater than 500.

In the manner of Chapter 9 we have

$$b_2 = \frac{m_{11}}{m_{20}}$$

$$\frac{\delta b_2}{b_2} = \frac{\delta m_{11}}{m_{11}} - \frac{\delta m_{20}}{m_{20}},$$

giving

$$\frac{\text{var } b_2}{b_2^2} = \frac{\text{var } m_{11}}{m_{11}^2} + \frac{\text{var } m_{20}}{m_{20}^2} - \frac{2 \text{ cov } (m_{11}, m_{20})}{m_{11}m_{20}} \quad . \quad (14.24)$$

Substituting from (9.16) and (9.17), and writing the sampling values m instead of the parent μ's, we have

$$\text{var } b_2 = \frac{b_2^2}{n}\left(\frac{m_{22}}{m_{11}^2} + \frac{m_{40}}{m_{20}^2} - \frac{2m_{31}}{m_{11}m_{20}}\right). \quad . \quad (14.25)$$

or, for the normal case, on using the values of Example 3.15,

$$\text{var } b_2 = \frac{b_2^2}{n}\left(\frac{1 - r^2}{r^2}\right.$$

$$\frac{1}{n}\frac{\text{var } y}{\text{var } x}(1 \quad . \quad (14.26)$$

Similarly

$$\text{var } b_1 = \frac{1}{n}\frac{\text{var } x}{\text{var } y}(1 - r^2). \quad . \quad (14.27)$$

To our order of approximation it is indifferent whether we write $1 - r^2$ or $1 - \rho^2$ on the right-hand side of these equations.

Example 14.5

In the data of Table 14.3 (height of fathers and height of sons) we find $r = +0{\cdot}51$, for $n = 1078$. From inspection of the table we see that the distribution is reasonably close to the normal type, and in this case n is large enough to justify the use of the standard error. We have then

$$\text{var } r = \frac{\{1 - (0{\cdot}51)^2\}^2}{1078}$$
$$= 0{\cdot}000{,}508.$$

Thus the standard error is about $0{\cdot}023$. The correlation is thus undoubtedly significant, if the data were obtained by random sampling. It is improbable that the parent correlation ρ lies outside the range $0{\cdot}51 \pm 0{\cdot}05$, and very improbable that it lies outside the range $0{\cdot}51 \pm 0{\cdot}075$.

Estimates of Correlation and Regression Coefficients in Normal Samples

14.11. In large-sample theory the sample values of the correlation and regression coefficients may be taken as estimates of the population values in the usual way. They can also be used in small-sample theory and it may, in fact, be shown that they are estimates giving maximum likelihood to samples from a bivariate normal population.

The joint probability of n sample values $(x_1, y_1) \ldots (x_n, y_n)$ from a bivariate normal population with means m_1 and m_2 is

$$dF = \frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n (1-\rho^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\Sigma\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho\Sigma\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \Sigma\left(\frac{y-m_2}{\sigma_2}\right)^2\right\}\right] dx_1\, dy_1 \ldots dx_n\, dy_n \quad (14.28)$$

The likelihood function may then be written

$$L \propto \frac{1}{\sigma_1^n \sigma_2^n (1-\rho^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2(1-\rho^2)}\{A - 2\rho B + C\}\right] \quad . \qquad (14.29)$$

and thus, for the maximisation of log L we have

$$\frac{1}{L}\frac{\partial L}{\partial m_1} = -\frac{1}{2(1-\rho^2)}\left\{-\frac{2}{\sigma_1^2}\Sigma(x-m_1) + \frac{2\rho}{\sigma_1\sigma_2}\Sigma(y-m_2)\right\} = 0$$

giving

$$\frac{1}{\sigma_1}\Sigma(x-m_1) - \frac{\rho}{\sigma_2}\Sigma(y-m_2) = 0 \quad . \quad . \qquad (14.30)$$

Similarly from $\frac{1}{L}\frac{\partial L}{\partial m_2} = 0$ we have

$$\frac{\rho}{\sigma_1}\Sigma(x-m_1) - \frac{1}{\sigma_2}\Sigma(y-m_2) = 0 \qquad (14.31)$$

Thus from (14.30) and (14.31), ρ not being unity in general,

$$\Sigma(x-m_1) = \Sigma(y-m_2) = 0$$

$$\left.\begin{aligned} m_1 &= \frac{1}{n}\Sigma(x) \\ m_2 &= \frac{1}{n}\Sigma(y) \end{aligned}\right\} \qquad (14.32)$$

so that our estimates of the means are the means of the sample.

We also find, equating $\frac{\partial}{\partial\sigma_1}\log L$, $\frac{\partial}{\partial\sigma_2}\log L$ and $\frac{\partial}{\partial\rho}\log L$ to zero and cancelling factors in σ_1, σ_2 and $(1-\rho^2)$ respectively,

$$\begin{aligned} -n + \frac{1}{1-\rho^2}(A - \rho B) &= 0 \\ -n + \frac{1}{1-\rho^2}(-\rho B + C) &= 0 \\ -n + \frac{1}{1-\rho^2}(A - 2\rho B + C) - \frac{B}{\rho} &= 0 \end{aligned} \qquad (14.33)$$

whence

$$\rho = \frac{B}{n}$$

$$1 = \frac{A}{n} = \frac{C}{n},$$

giving for the estimates of σ_1^2, σ_2^2 and ρ

$$\left.\begin{aligned}\sigma_1^2 &= \frac{1}{n}\Sigma(x - m_1)^2 \\ \sigma_2^2 &= \frac{1}{n}\Sigma(y - m_2)^2 \\ \rho &= \frac{\Sigma(x - m_1)(y - m_2)}{\{\Sigma(x - m_1)^2\ \Sigma(y - m_2)^2\}^{\frac{1}{2}}}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (14.34)$$

which, on substituting the estimates of m_1 and m_2 given by (14.32), become the sample variances and correlation coefficient.

Distribution of Sample Means, Variances and Covariance in Normal Samples

14.12. In accordance with the result of the previous section we will take our estimates of the parameters m_1, m_2, σ_1^2, σ_2^2 and ρ to be the corresponding sample values $\bar{x}$, $\bar{y}$, s_1^2, s_2^2 and r. The joint distribution of the sample values is given by (14.28) and it is remarkable that the exponent in that expression can be expressed solely in terms of the five parameters and their estimates. We have, in fact,

$$\Sigma\left(\frac{x - m_1}{\sigma_1}\right)^2 - 2\rho\Sigma\frac{(x - m_1)(y - m_2)}{\sigma_1\sigma_2} + \Sigma\left(\frac{y - m_2}{\sigma_2}\right)^2$$

$$= \Sigma\frac{(x - \bar{x} + \bar{x} - m_1)^2}{\sigma_1^2} + \text{two similar terms}$$

$$= \Sigma\frac{(x - \bar{x})^2}{\sigma_1^2} + \Sigma\frac{(\bar{x} - m_1)^2}{\sigma_1^2} + \text{four similar terms,}$$

the product terms vanishing because $\Sigma(x - \bar{x}) = \Sigma(y - \bar{y}) = 0$,

$$= n\left\{\frac{s_1^2 + (\bar{x} - m_1)^2}{\sigma_1^2}\right\} + \text{two similar terms}$$

$$= n\left\{\frac{(\bar{x} - m_1)^2}{\sigma_1^2} - 2\rho\frac{(\bar{x} - m_1)(\bar{y} - m_2)}{\sigma_1\sigma_2} + \frac{(\bar{y} - m_2)^2}{\sigma_2^2}\right\}$$

$$+ n\left\{\frac{s_1^2}{\sigma_1^2} - \frac{2\rho r s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\} \quad . \quad (14.35)$$

We proceed to find the joint distribution of the five statistics, and to do so require to express the frequency element (14.28) in terms of them. The non-differential part of that element is given by the exponential of (14.35). It remains to express in the requisite form the volume element $dx_1 \ldots dx_n\, dy_1 \ldots dy_n$.

Generalising the geometrical approach of Chapter 10, we may imagine a sample space of $2n$ dimensions, n for x and n for y. The sample point may vary in the x-space and the y-space, but not independently so. In fact, if P represents the point $(x_1 \ldots x_n)$ in the x-space and Q the point $(y_1 \ldots y_n)$ in the y-space, and if O_1, O_2 are the points $(\bar{x} \ldots \bar{x})$, $(\bar{y} \ldots \bar{y})$, then for any given r we have

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{ns_1s_2} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\{\Sigma(x - \bar{x})^2\,\Sigma(y - \bar{y})^2\}^{\frac{1}{2}}}$$

and thus r is the cosine of the angle, say θ, between PO_1 and QO_2, so that if P and r are

fixed Q varies on the cone in the y-space obtained by rotating O_2Q such that the angle made with O_1P is constant.

The element in the x-space is proportional to $s_1^{n-2}\,ds_1\,d\bar{x}_1$ as was seen in Example 10.5. For given r, $\bar{y}$ and s_2 the point Q varies on the zone of the hypersphere of radius $s_2\sqrt{n}$, centre $\bar{y}$ and $(n-1)$ dimensions. This zone has radius $s_2\sqrt{n}\sin\theta = s_2\sqrt{n}\,(1-r^2)^{\frac{1}{2}}$ and width $s_2\sqrt{n}\,d\theta = \dfrac{s_2\sqrt{n}\,dr}{(1-r^2)^{\frac{1}{2}}}$ and thus its content is proportional to $\{s_2\sqrt{n}(1-r^2)^{\frac{1}{2}}\}^{n-3}$ $\dfrac{s_2\sqrt{n}\,dr}{(1-r^2)^{\frac{1}{2}}}$, that is, to $s_2^{n-2}(1-r^2)^{\frac{n-4}{2}}$.

Thus the volume element may be written

$$dv \propto s_1^{n-2}\,ds_1\,d\bar{x} \quad s_2^{n-2}\,ds_2\,d\bar{y}\,(1-r^2)^{\frac{n-4}{2}}\,dr$$

$$\propto s_1^{n-2}\,s_2^{n-2}\,ds_1\,ds_2\,(1-r^2)^{\frac{n-4}{2}}\,dr\,d\bar{x}\,d\bar{y} \quad . \quad . \quad . \quad (14.36)$$

and the joint frequency element of the five variables is then proportional to

$$\exp -\frac{n}{2(1-\rho^2)}\left[\left\{\frac{(\bar{x}-m_1)^2}{\sigma_1^2} - 2\rho\frac{(\bar{x}-m_1)(\bar{y}-m_2)}{\sigma_1\sigma_2} + \frac{(\bar{y}-m_2)^2}{\sigma_2^2}\right\}\right.$$
$$\left. + n\left\{\frac{s_1^2}{\sigma_1^2} - 2\rho r\frac{s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right]dv \quad (14.37)$$

This fundamental result is due to R. A. Fisher (1915).

14.13. One important property of (14.37) may be remarked. The distribution may be factorised into two parts, one containing only $\bar{x}$ and $\bar{y}$ and the other only s_1, s_2 and r, namely (except for constants)

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{(\bar{x}-m_1)^2}{\sigma_1^2} - 2\rho\frac{(\bar{x}-m_1)(\bar{y}-m_2)}{\sigma_1\sigma_2} + \frac{(\bar{y}-m_2)^2}{\sigma_2^2}\right\}\right]d\bar{x}\,d\bar{y} \quad . \quad (14.38)$$

and

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2} - \frac{2\rho r s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right]s_1^{n-2}\,s_2^{n-2}(1-r^2)^{\frac{n-4}{2}}\,ds_1\,ds_2\,dr \quad . \quad (14.39)$$

Thus we see that in normal samples the distribution of means is entirely independent of that of variances and covariance.

Before leaving (14.38) we may also note that the means are themselves distributed in the bivariate normal form, with mean $(\bar{x}) = m_1$, mean $(\bar{y}) = m_2$, var $(\bar{x}) = \dfrac{\sigma_1^2}{n}$, var $(\bar{y}) = \dfrac{\sigma_2^2}{n}$ (all of which results are already familiar), and

$$\text{cov}\,(\bar{x}, \bar{y}) = \frac{\sigma_1\sigma_2\rho}{n}, \quad . \quad (14.40)$$

so that the correlation between $\bar{x}$ and $\bar{y}$ is ρ, the correlation in the parent population.

14.14. We may now use (14.37) to obtain the distribution of the correlation coefficient, namely, by integrating with respect to s_1 and s_2 from 0 to ∞. Let us first of all evaluate the constant to be attached to it from the consideration that $\int dF = 1$.

Make the variate transformation

$$\left.\begin{aligned} a &= \frac{s_1^2}{\sigma_1^2}\cdot\frac{n}{2(1-\rho^2)} \\ b &= \frac{rs_1s_2}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} \\ c &= \frac{s_2^2}{\sigma_2^2}\cdot\frac{n}{2(1-\rho^2)} \end{aligned}\right\} \qquad (14.41)$$

We have for the Jacobian of the transformation

$$\frac{\partial(a,b,c)}{\partial(s_1,r,s_2)} = \begin{vmatrix} \frac{2s_1}{\sigma_1^2}\cdot\frac{n}{2(1-\rho^2)} & 0 & 0 \\ \frac{rs_2}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} & \frac{s_1s_2}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} & \frac{rs_1}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} \\ & 0 & \frac{2s_2}{\sigma_2^2}\cdot\frac{n}{2(1-\rho^2)} \end{vmatrix}$$

$$= \frac{s_1^2s_2^2\,n^3}{2\sigma_1^3\sigma_2^3(1-\rho^2)^3} = \frac{2acn}{\sigma_1\sigma_2(1-\rho^2)}$$

and also the relation

$$r^2 = \frac{b^2}{ac}.$$

The integral then becomes

$$\int \exp[-a+2\rho b-c].\left\{\frac{2a\sigma_1^2(1-\rho^2)}{n}\right\}^{\frac{n-2}{2}}\left\{\frac{2c\sigma_2^2(1-\rho^2)}{n}\right\}^{\frac{n-2}{2}}$$
$$\times\left(1-\frac{b^2}{ac}\right)^{\frac{n-4}{2}}.\frac{\sigma_1\sigma_2(1-\rho^2)}{2acn}\,da\,db\,dc$$

$$= \frac{2^{n-3}\,\sigma_1{}^{n-1}\,\sigma_2{}^{n-1}\,(1-\rho^2)^{n-1}}{n^{n-1}}\int \exp[-a+2\rho b-c]\,(ac-b^2)^{\frac{n-4}{2}}\,da\,db\,dc \qquad (14.42)$$

where the limits of a and c are 0 to ∞ and those of b are $-\infty$ to $+\infty$. This integral may be evaluated in terms of the Γ-function. Putting $\xi = a - \frac{b^2}{c}$ we find

$$\int \exp(-\xi)\exp\left(+2\rho b - c - \frac{b^2}{c}\right)\xi^{\frac{n-4}{2}}\,c^{\frac{n-4}{2}}\,d\xi\,db\,dc$$

$$= \Gamma\left(\frac{n-2}{2}\right)\int \exp\left[-\frac{1}{c}\{(b-\rho c)^2+(1-\rho^2)c^2\}\right]c^{\frac{n-4}{2}}\,db\,dc$$

$$= \Gamma\left(\frac{n-2}{2}\right)\sqrt{\pi}\int \exp\{-(1-\rho^2)c\}c^{\frac{n-3}{2}}\,dc$$

$$= \sqrt{\pi}\,\frac{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{(1-\rho^2)^{\frac{n-1}{2}}}$$

$$= \frac{\pi\,\Gamma(n-2)}{2^{n-3}(1-\rho^2)^{\frac{n-1}{2}}} \qquad (14.43)$$

Collecting up the terms from (14.42) and (14.43) we find for the joint distribution of s_1, s_2 and r

$$dF = \frac{n^{n-1}}{\pi\sigma_1^{n-1}\,\sigma_2^{n-1}\,(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma(n-2)}\exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2}-\frac{2\rho r s_1 s_2}{\sigma_1\sigma_2}+\frac{s_2^2}{\sigma_2^2}\right\}\right]$$
$$s_1^{n-2}\,s_2^{n-2}(1-r^2)^{\frac{n-4}{2}}\,ds_1\,ds_2\,dr \quad (14.44)$$

Now put

$$\zeta = \frac{s_1 s_2}{\sigma_1\sigma_2},\quad z = \log\frac{\sigma_2 s_1}{\sigma_1 s_2},\quad r = r.$$

We find, for the Jacobian of the transformation

$$\frac{\partial(\zeta, z, r)}{\partial(s_1, s_2, r)} = \begin{vmatrix} \dfrac{s_2}{\sigma_1\sigma_2} & \dfrac{s_1}{\sigma_1\sigma_2} & 0 \\ \dfrac{1}{s_1} & -\dfrac{1}{s_2} & 0 \\ 0 & 0 & 1 \end{vmatrix}$$
$$\frac{2}{\sigma_1\sigma_2}$$

The exponent in (14.44) becomes

$$\exp\left[-\frac{n}{2(1-\rho^2)}\{\zeta e^{z}-2\rho r\zeta+\zeta e^{-z}\}\right]$$

and after a little reduction the distribution becomes

$$dF = \frac{n^{n-1}}{\pi(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma(n-2)}\exp\left[-\frac{n}{(1-\rho^2)}\zeta(\cosh z-\rho r)\right]\zeta^{n-2}\,d\zeta\,dz(1-r^2)^{\frac{n-4}{2}}\,dr.$$

On integration with respect to ζ we have

$$dF = \frac{(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma(n-1)}{\pi\,\Gamma(n-2)}\,\frac{(1-r^2)^{\frac{n-4}{2}}}{(\cosh z-\rho r)^{n-1}}\,dz\,dr.$$

Putting $-\rho r = \cos\theta$ we have, since

$$\int_0^\infty \frac{dz}{\cosh z+\cos\theta} = \frac{\theta}{\sin\theta},$$

$$dF = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)}(1-r^2)^{\frac{n-4}{2}}\,\frac{d^{n-2}}{d(-\cos\theta)^{n-2}}\left(\frac{\theta}{\sin\theta}\right)dr$$
$$= \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)}(1-r^2)^{\frac{n-4}{2}}\,\frac{d^{n-2}}{d(r\rho)^{n-2}}\left\{\frac{\cos^{-1}(-\rho r)}{\sqrt{(1-\rho^2r^2)}}\right\}dr \quad . \quad (14.45)$$

This is as simple a form as can be given in terms of elementary functions.

14.15. In the particular case $\rho = 0$, (14.45) reduces to

$$dF = \frac{1}{B\left(\frac{n-2}{2},\frac{1}{2}\right)}(1-r^2)^{\frac{n-4}{2}}\,dr \quad . \quad . \quad . \quad (14.46)$$

a form surmised by "Student" in 1908. This distribution provides a test of the hypothesis that an observed r arose from an uncorrelated normal population. Its distribution function may be obtained from incomplete B-functions, or more conveniently by putting

$$t = \frac{r}{\sqrt{(1-r^2)}}\sqrt{(n-2)}\ . \tag{14.47}$$

which transforms (14.46) to

$$dF = \frac{dt}{\sqrt{(n-2)}\,B\left(\frac{n-2}{2}, \frac{1}{2}\right)\left(1+\frac{t^2}{n-2}\right)^{\frac{n-1}{2}}}\ . \tag{14.48}$$

The integral of this function has been tabulated and is given as Appendix Table 3. Fisher and Yates have also tabulated some of the significance points of t and of r itself, i.e. the values of r (for various n) for which the distribution function takes specified values.

14.16. The general distribution (14.45) has been studied in some detail, but lack of space prevents the inclusion of the extensive analysis involved. We will here indicate only the more important features of the results.

First, as to the shape of the frequency curves. When $n = 2$ the distribution becomes

$$dF = \frac{(1-\rho^2)\cos^{-1}(-\rho r)}{\pi(1-r^2)\sqrt{(1-\rho^2 r^2)}}\,dr$$

and the frequency curve may be written

$$y = y_0\frac{\theta}{(1-r^2)\sin\theta} \qquad \theta = \cos^{-1}(-\rho r). \tag{14.49}$$

For $r = \pm 1$ the ordinate is infinite and the distribution will be found to be U-shaped.

When $n = 3$ we find

$$y = y_0\left\{\frac{1}{\sin^2\theta} + \frac{\theta\cos\theta}{\sin^3\theta}\right\}\frac{1}{(1-r^2)^{\frac{1}{2}}}\ . \tag{14.50}$$

again a U-shaped distribution. For $n = 4$,

$$y = \frac{y_0}{\sin^3\theta}\{\theta - 3\cot\theta + 3\theta\cot^2\theta\}. \tag{14.51}$$

If $\rho = 0$ this reduces to the rectangular form $y = \frac{1}{2}$. In other cases the curve is J-shaped, increasing from a minimum at $r = -1$ to a maximum (but not an infinite maximum) at $r = +1$.

For $n > 4$ the frequency curves are unimodal and tend to normality with large n, though slowly. Some interesting photographs of models of these curves are given in the "Co-operative Study" (1917).

14.17. The moments of the distribution are expressible in terms of hypergeometric functions. Returning to (14.44) let us write

$$\alpha_1^2 = \frac{\sigma_1^2(1-\rho^2)}{n}, \qquad \alpha_2^2 = \frac{\sigma_2^2(1-\rho^2)}{n}.$$

After a little rearrangement the distribution becomes

$$dF = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)} \exp\left\{-\frac{1}{2}\frac{s_1^2}{\alpha_1^2} - \frac{1}{2}\frac{s_2^2}{\alpha_2^2} + \rho r\frac{s_1 s_2}{\alpha_1\alpha_2}\right\}$$

$$\left(\frac{s_1}{\alpha_1}\right)^{n-2}\left(\frac{s_2}{\alpha_2}\right)^{n-2}(1-r^2)^{\frac{n-4}{2}}\left(\frac{ds_1}{\alpha_1}\right)\left(\frac{ds_2}{\alpha_2}\right)dr \quad . \quad . \quad (14.52)$$

Putting $u_1 = \frac{s_1}{\alpha_1}$, $u_2 = \frac{s_2}{\alpha_2}$ and expanding the term in $\exp\left(\rho r\frac{s_1 s_2}{\alpha_1\alpha_2}\right)$ we have

$$dF = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)} \exp\left(-\tfrac{1}{2}u_1^2 - \tfrac{1}{2}u_2^2\right) u_1^{n-2}\, u_2^{n-2}(1-r^2)^{\frac{n-4}{2}}$$

$$\times \sum_{j=0}^{\infty} \frac{(\rho r u_1 u_2)^j}{j!}\, du_1\, du_2\, dr \quad . \qquad (14.53)$$

Integrating for u_2 from 0 to ∞ we find for the distribution of u_1 and r

$$dF = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)} \exp\left(-\tfrac{1}{2}u_1^2\right)u_1^{n-2}(1-r^2)^{\frac{n-4}{2}}\, du_1\, dr\, \Sigma\frac{(\rho r u_1)^j}{j!}\,\Gamma\left(\frac{n-1+j}{2}\right).2^{\frac{n+j-3}{2}}.$$

Multiplying by r and integrating from -1 to $+1$ we find

$$\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)} \exp\left(-\tfrac{1}{2}u_1^2\right) u_1^{n-2}\, du_1 \times \sum_{j=0}^{\infty}\frac{(\rho u_1)^{2j+1}}{(2j+1)!}\,\Gamma\left(\frac{n+2j}{2}\right) B\left(\frac{n-2}{2}, \frac{2j+3}{2}\right).2^{\frac{n+2j-2}{2}}$$

and finally, integrating with respect to u_1 we obtain

$$\mu_1(r) = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\Gamma(n-2)} \sum_{j=0}^{\infty} \frac{\rho^{2j+1}}{(2j+1)!}\,\Gamma\left(\frac{n+2j}{2}\right) B\left(\frac{n-2}{2}, \frac{2j+3}{2}\right)\Gamma\left(\frac{n+2j}{2}\right) 2^{n+2j-2} \quad (14.54)$$

Substituting for the B-function in terms of Γ-functions and remembering that

$$\Gamma(x)\,\Gamma(x+\tfrac{1}{2}) = \frac{\pi^{\frac{1}{2}}}{2^{2x-1}}\,\Gamma(2x),$$

we find

$$\mu_1'(r) = \frac{\rho(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma^2\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n+1}{2}\right)}\left\{1 + \frac{\frac{1}{2}n.\frac{1}{2}n}{\frac{1}{2}(n+1)}\frac{\rho^2}{2!} + \frac{\frac{1}{2}n(\frac{1}{2}n+1).\frac{1}{2}n(\frac{1}{2}n+1)}{\frac{1}{2}(n+1).\frac{1}{2}(n+3)}\frac{\rho^4}{4!} + \ldots\right.$$

$$= \frac{\rho(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma^2\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n+1}{2}\right)}\, F\left(\tfrac{1}{2}n, \tfrac{1}{2}n, \tfrac{1}{2}(n+1), \rho^2\right)$$

and since $F(\alpha, \beta, \gamma, x) = (1-x)^{\gamma-\alpha-\beta}F(\gamma-\alpha, \gamma-\beta, \gamma, x)$

$$\mu_1'(r) = \frac{\rho\,\Gamma^2\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n+1}{2}\right)}\, F\left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}(n+1), \rho^2\right) \quad . \quad . \quad . \quad (14.55)$$

In a similar way it may be shown that

$$\mu_2'(r) = 1 - (1-\rho^2)\,\frac{n-2}{n-1}\,F(1, 1, \tfrac{1}{2}(n+1), \rho^2) \quad . \quad . \quad . \quad (14.56)$$

These series converge fairly rapidly for moderate or large n.

14.18. The ordinates and distribution function of the correlation coefficient are not expressible in terms of simple mathematical functions. They have, however, been tabulated by David (1938) for values of $n = 2(1)25$, 50, 100, 200 and 400; for $\rho = 0{\cdot}0(0{\cdot}1)0{\cdot}9$; and for $r = -1{\cdot}00(0{\cdot}05) + 1{\cdot}00$, with finer intervals in places.

For many practical purposes it is sufficient to use a transformation of the distribution due to Fisher (1921). Putting

$$\left.\begin{aligned} r &= \tanh z, \quad z = \tfrac{1}{2}\log\frac{1+r}{1-r} \\ \rho &= \tanh \zeta, \quad \zeta = \tfrac{1}{2}\log\frac{1+\rho}{1-\rho} \end{aligned}\right\} \quad . \quad (14.57)$$

we may expand the frequency function of r in powers of $z - \zeta$, $= x$ say, and inverse powers of n. Fisher gives the following expansion:

$$\begin{aligned} f = \frac{n-2}{\sqrt{\{2\pi(n-1)\}}}\,e^{-\frac{n-1}{2}x^2}\Bigg\{1 &+ \tfrac{1}{2}\rho x + \left(\frac{2+\rho}{8(n-1)} + \frac{4-\rho^2}{8}x^2 + \frac{n-1}{12}x^4\right) \\ &+ \rho x\left(\frac{4-\rho^2}{16(n-1)} + \frac{4+3\rho^2}{48}x^2 + \frac{n-1}{24}x^4\right) + \left(\frac{4+12\rho^2+9\rho^4}{128(n-1)^2} + \frac{8-2\rho^2+3\rho^4}{64(n-1)}\right. \\ &\left. + \frac{4\rho^2-5\rho^4}{128}x^4 + \frac{28-15\rho^2}{1440}x^6(n-1)\right) + \frac{(n-1)^2}{288}x^8 + \ldots \end{aligned} \quad . \quad (14.58)$$

Taking moments about $x = 0$ we find, on transferring to the mean,

$$\mu_1' = \frac{\rho}{2(n-1)}\left\{1 + \frac{1+\rho^2}{8(n-1)} + \ldots\right\} \quad . \quad . \quad . \quad . \quad (14.59)$$

$$\mu_2 = \frac{1}{n-1}\left\{1 + \frac{4-\rho^2}{2(n-1)} + \frac{176-21\rho^2-21\rho^4}{48(n-1)^2} + \ldots\right\} \quad . \quad (14.60)$$

$$\mu_3 = \frac{\rho(\rho^2 - \frac{9}{16})}{(n-1)^3} \quad . \quad (14.61)$$

$$\frac{1}{(n-1)^2}\left\{\ldots \frac{224-48\rho^2-3\rho^4}{16(n-1)} + \frac{1472-228\rho^2-141\rho^4-3\rho^6}{32(n-1)^2} \ldots\right\} \quad . \quad (14.62)$$

The remarkable thing about the transformation is that the distribution of r, which is very skew, becomes the distribution of $z - \zeta$, which is nearly symmetrical. In fact,

$$\gamma_1 = \frac{\rho}{(n-1)^{\frac{3}{2}}}(\rho^2 - \tfrac{9}{16}) \quad . \quad (14.63)$$

$$\gamma_2 = \frac{32-3\rho^4}{16(n-1)} + \ldots \quad . \quad (14.64)$$

Thus we may take $z - \zeta$ to be approximately normally distributed with mean and variance given by (14.59) and (14.60). As a slightly rougher approximation we may take

$$\mu_1'(z - \zeta) = \frac{\rho}{2(n-1)} \tag{14.65}$$

$$\operatorname{var}(z - \zeta) = \frac{}{n-1} + \frac{}{2(n-1)^2}$$

which is approximately equal for small ρ to

$$\frac{1}{n-1} + \frac{}{(n-1)^2}$$

$$\frac{1}{n-3} \text{ approximately .} \tag{14.66}$$

When n is moderate we may take a still rougher approximation by assuming $z - \zeta$ to be normally distributed about zero mean with variance $\frac{1}{n-3}$. Some comparisons of the various approximations are given in the introduction to David's tables, and it appears that for $n > 50$ the forms (14.65) and (14.66) are adequate. The approximation given by (14.59) and (14.60) appears to hold satisfactorily for values of n as low as 11.

14.19. Except in the case of the normal parent very little exact knowledge is available about the sampling distribution of the correlation coefficient. There is, however, some empirical evidence to justify the use of the above results when the population does not differ very much from the normal. E. S. Pearson (1931), in dealing with some experimental results, concluded that "the results suggest that the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of r." The subject does not seem to have been investigated mathematically except in special cases.

Example 14.6

In Example 14.3 we obtained for the correlation coefficient between wheat and potatoes a value of 0·2182. Suppose we regard the 48 counties as giving a random sample of the yields of wheat and potatoes, either for a wider area or for an extended period of years. The question then is, can such a value have arisen by chance from a population in which the yields of wheat and potatoes are uncorrelated?

From prior knowledge of crop yields we can assume with some confidence approximate normality in the parent population. Let us then test the hypothesis that the correlation in this population is zero.

We have

$$z = \tfrac{1}{2}\log_e \frac{1+r}{1-r} = \tfrac{1}{2}\log_e \frac{1{\cdot}2189}{0{\cdot}7811} \qquad 0{\cdot}2225$$

$$\zeta = 0$$

$$\frac{1}{\sqrt{(n-3)}} \qquad \frac{1}{\sqrt{(45)}} \qquad 0{\cdot}1491.$$

The deviation $z - \zeta$ is thus 0·2225, or about 1·49 times the standard error. This is not very improbable and the observed correlation may thus be accidental.

Example 14.7

In a sample of 50 a correlation coefficient is found to be $+ 0{\cdot}5$. What is the probability that a value equal to or *less* than this should have been obtained from a normal population in which the correlation is $+ 0{\cdot}7$?

The exact value, from David's table, is, to five decimal places, 0·01289. Let us first of all take the approximation which assumes $z - \zeta$ to be distributed about zero mean with variance $\frac{1}{(n-3)}$. We have

$$z = \tfrac{1}{2}\log\frac{1+r}{1-r} = 0{\cdot}5493$$

$$\zeta = \tfrac{1}{2}\log\frac{1+\rho}{1-\rho} = 0{\cdot}8673$$

$$\frac{1}{\sqrt{(n-3)}} = 0{\cdot}1459, \quad z - \zeta = -\,0{\cdot}3180.$$

The deviation is thus 2·18 times the standard error, and the required probability, from the table of the normal integral, 0·0146 approximately, compared with the true value of 0·0129. The approximate test is not quite stringent enough.

Let us then take $z - \zeta$ to be distributed normally about mean $\frac{\rho}{2(n-1)} = 0{\cdot}00714$. The deviation is then $-\ 0{\cdot}3251$, or 2·23 times the standard error, giving a probability of about 0·0129, almost the exact value.

Example 14.8

In a sample of n_1 there is observed a correlation of r_1 and in a second sample of n_2 a correlation of r_2. Are the sample values r_1 and r_2 compatible with the hypothesis that the samples arose from the same population?

Suppose the hypothesis were true, and that ρ is the correlation coefficient in the population. Then if $z_1 = \tanh^{-1} r_1$, $z_2 = \tanh^{-1} r_2$, $\zeta = \tanh^{-1}\rho$, we know that if the population were normal, $z_1 - \zeta$ will be distributed approximately with variance $\frac{1}{n_1 - 3}$ and $z_2 - \zeta$ with variance $\frac{1}{n_1 - 3}$. Thus the difference $z_1 - z_2 = (z_1 - \zeta) - (z_2 - \zeta)$ is distributed approximately normally with variance

$$\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3},$$

and this will provide a test of the hypothesis.

Distribution of Regression Coefficients in Normal Samples

14.20. Turning again to equation (14.44) we have, substituting $b_2 = \frac{rs_2}{s_1}$, the joint frequency-distribution of s_1, s_2 and b_2

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2} - \frac{2\rho s_1^2 b_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right] s_1^{\,n-1}\, s_2^{\,n-3}\left(1 - \frac{s_1^2 b_2^2}{s_2^2}\right)^{\frac{n-4}{2}} ds_1\, ds_2\, db_2 \quad (14.67)$$

Integration with respect to s_2 gives for the distribution of s_1 and b_2

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2} - \frac{2\rho s_1^2 b_2}{\sigma_1\sigma_2} + \frac{s_1^2 b_2^2}{\sigma_2^2}\right\}\right] s_1^{\,n-1}\, ds_1\, db_2.$$

A further integration with respect to s_1 gives for the distribution of b_2

$$dF \propto \frac{db_2}{\left(1 - \frac{2\rho\sigma_1}{\sigma_2}b_2 + \frac{\sigma_1^2}{\sigma_2^2}b_2^2\right)^{\frac{n}{2}}}$$

or, on evaluation of the constant,

$$dF = \frac{\Gamma\left(\frac{n}{2}\right)\sigma_1^{\,n-1}(1-\rho^2)^{\frac{n-1}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\sigma_2^{\,n-2}}\,\frac{db_2}{\left\{\frac{\sigma_2^2}{\sigma_1^2}(1-\rho^2) + \left(b_2 - \frac{\rho\sigma_2}{\sigma_1}\right)^2\right\}^{\frac{n}{2}}} \quad . \quad . \quad (14.68)$$

The distribution of the regression coefficient b_1 is obtainable by interchanging the suffixes 1 and 2.

The form (14.68) is a Pearson Type VII distribution, symmetrical about the point $b_2 = \frac{\rho\sigma_2}{\sigma_1}$, the population regression coefficient. It tends to normality fairly rapidly, and the use of the standard error for regressions is therefore valid for lower values of n than in the case of the correlation coefficient. For small samples, however, (14.68) is not of much use since it depends on the unknown quantities σ_1, σ_2 and ρ, i.e. the population variances and covariance.

14.21. It is possible to find statistics other than b_1 and b_2 which will provide a test of the regressions. Write

$$u = \frac{(b_2 - \beta_2)s_1}{(s_2^2 - b_2^2 s_1^2)^{\frac{1}{2}}}. \quad . \quad . \quad . \quad . \quad . \quad (14.69)$$

We now return to the distribution of the quantities a, b, c of equation (14.41), namely,

$$dF \propto \exp\left[-a + 2\rho b - c\right](ac - b^2)^{\frac{n-4}{2}}\, da\, db\, dc \quad . \quad . \quad . \quad (14.70)$$

We have from (14.69)

$$u = \frac{b - \rho a}{\sqrt{(ac - b^2)}},$$

and on substituting for c in (14.70) we have, after a little reduction,

$$dF \propto \frac{\exp\left[-a\left(1 + \frac{\rho^2}{u^2}\right)\right] da\, du}{au^{n-1}}\,.\exp\left[\frac{u^2+1}{u^2}\left(2\rho b - \frac{b^2}{a}\right)\right](b - \rho a)^{n-2}\, db.$$

The integral of the second part on the right for b will be found to give a factor proportional to

$$\exp\frac{a\rho^2(u^2+1)}{u^2}\cdot\left(\frac{u^2}{u^2+1}\right)^{n-1}$$

and hence for the distribution of a and u we find

$$dF \propto \frac{a^{\frac{n-1}{2}}\exp(-a + \rho^2 a)\, da}{a}\cdot\frac{du}{(1+u^2)^{\frac{n-1}{2}}}.$$

Hence the distributions of a and u are independent, and for that of u we have

$$dF \propto \frac{du}{(1+u^2)^{\frac{n-1}{2}}}. \qquad (14.71)$$

This distribution does not contain any of the parent parameters. If we put

$$t = u\sqrt{(n-2)} = \frac{(b_2-\beta_2)s_1\sqrt{(n-2)}}{(s_2^2-b_2^2s_1^2)^{\frac{1}{2}}}, \qquad (14.72)$$

then t is distributed in "Student's" form

$$dF \propto \frac{dt}{\left(1+\frac{t^2}{n-2}\right)^{\frac{n-1}{2}}} \qquad (14.73)$$

and may be tested accordingly.

Example 14.9

In Example 14.3 we found for the regression of Y (potato yield) on X (wheat yield)

$$(Y - 6{\cdot}065) = 0{\cdot}0783\,(X - 15{\cdot}791).$$

The regression coefficient is small. Could it have arisen from a population in which there is no correlation, i.e. in which $\beta_2 = 0$?

From Example 14.3 we have

$$b_2 = 0{\cdot}0783\ \sqrt{(n-2)} = 6{\cdot}7823,\ s_1^2 = 4{\cdot}1749,\ s_2^2 = 0{\cdot}5340.$$

Hence from (14.72)

$$t = \frac{b_2s_1\sqrt{(n}}{\sqrt{(s_2^2-s_1^2b_2^2)}} = 2{\cdot}06.$$

Appendix Table 3 does not carry us as far as $\nu = 46$. From the Fisher–Yates tables, however, we have the following values of t for $P = 0{\cdot}05$:

$$\nu = 40 \quad t = 2{\cdot}021\ ; \qquad \nu = 60 \quad t = 2{\cdot}000,$$

and for $P = 0{\cdot}02$:

$$\nu = 40 \quad t = 2{\cdot}423\ ; \qquad \nu = 60 \quad t = 2{\cdot}390.$$

Thus in our case P evidently lies between 0·05 and 0·02, and the regression may not be significant, i.e. the two variates may be independent. This confirms the conclusion reached in Example 14.6 from consideration of the correlation coefficient.

14.22. Up to this point we have considered the correlation coefficient mainly as a measure of the relationship between two variates, and this is the standpoint which will mainly concern us in this and the succeeding chapter. We may, however, turn for a time to a consideration of the regression equations, which have an importance of their own. Assuming that the regression is approximately linear, we have two equations

$$\left.\begin{aligned} X-\bar{x} &= \beta_1(Y-\bar{y}) \\ Y-\bar{y} &= \beta_2(Y-\bar{x}) \end{aligned}\right\} \qquad (14.74)$$

expressing the relations between the means of variate arrays and the variate-values determining those arrays. A problem which frequently presents itself in practice is the following: given a member of the population exhibiting a variate-value x, what is its y-value? Evidently there is in general no unique answer to this question. For any given x there will be an

array of y's, any one of which might be exhibited by the member under consideration. But in the absence of any special knowledge it is reasonable to take as the best estimate of y the mean of this array. If the population is normal the mean will be the modal value, and if it is approximately normal the mean will be a reasonable estimate, the greater part of the population values lying distributed within a range of two or three times the standard deviation of the array.

In fact, the question as put is too restrictive. There is no unique value of y corresponding to a given x, and we are entitled to enquire only after the distribution of y's or their principal characteristics.

Now the mean required is given by the regression equation, and hence that equation may be used to *estimate* the y-value corresponding to a given x. If at the same time the variance of the y-array can be determined, the probable limits of error of the estimate may also be assigned. This is particularly easy for normal populations because, as we have seen (**14.8**), the variance of all x-arrays is $\sigma_1^2(1 - \rho^2)$ and that of the y-arrays $\sigma_2^2(1 - \rho^2)$. As usual in large samples, we can use the sample values to calculate these variances; or we may take the variance of the array direct from observation.

Example 14.10

In Example 14.1 we found for the regression equations, in the units there employed,

$$X - 0{\cdot}7706 = -\,1{\cdot}417\ (Y + 0{\cdot}2095)$$
$$Y + 0{\cdot}2095 = -\,0{\cdot}2658(X - 0{\cdot}7706).$$

Suppose we require to estimate the highest audible pitch for a man 34 years of age. In our units this corresponds to an x-value of $\frac{1}{3}(34 - 22) = 4$. Our estimate of y is then

$$\begin{aligned} &-\,0{\cdot}2095 - 0{\cdot}2658(4 - 0{\cdot}7706) \\ &= -\,1{\cdot}0679 \text{ units.} \end{aligned}$$

This corresponds, in vibrations per second, to

$$\begin{aligned} &19{,}995 - (1{\cdot}0679) \times 2000 \\ &= 17{,}900 \text{ vibrations.} \end{aligned}$$

The variance of the estimate is $s_1^2(1 - r^2)$

$$\begin{aligned} &= 13{\cdot}3482\{1 - (0{\cdot}6136)^2\} \text{ thousands}^2 \\ &= 8{\cdot}322 \text{ thousands}^2, \end{aligned}$$

so that the standard error is $\sqrt{8{\cdot}322} = 2{\cdot}9$ units $= 5{\cdot}8$ thousand vibrations. The estimate is evidently not very accurate, for the value of y can vary within two or three times this range without very great improbability.

If the problem had been set in the reverse form: what is the age corresponding to a vibration of 17·9 thousands, we should have

$$\begin{aligned} X &= 0{\cdot}7706 - 1{\cdot}417(-\,1{\cdot}0679 + 0{\cdot}2095) \\ &= 1{\cdot}99 \text{ units} \\ &= 27{\cdot}98 \text{ years.} \end{aligned}$$

This is not very close to 34 years, the age from which we started; and in general, if ξ is the estimate of x, given $y = \eta$, η is not the estimate of y, given $x = \xi$. We have a right to expect such a concordance only when r is near unity or when ξ and η are near the means of the distribution, where the regression lines intersect.

The Correlation Ratios

14.23. For any bivariate distribution we have, if $\bar{x}_p$ is the mean of the pth x-array and $\bar{x}$ the mean of the whole,

$$\Sigma(x - \bar{x})^2 = \Sigma(x - \bar{x}_p + \bar{x}_p - \bar{x})^2$$
$$= \Sigma(x - \bar{x}_p)^2 + \Sigma(\bar{x}_p - \bar{x})^2, \quad . \quad . \quad . \quad (14.75)$$

the product term $2\Sigma(x - \bar{x}_p)(\bar{x}_p - x)$ vanishing because $\bar{x}_p - \bar{x}$ is constant for any given array.

The correlation ratio of x on y, η_{xy}, is defined by

$$\eta^2_{xy} = \frac{\Sigma(\bar{x}_p - \bar{x})^2}{\Sigma(x - \bar{x})^2} \qquad (14.76)$$

and similarly that of y on x, η_{yx}, by

$$\eta^2_{yx} = \frac{\Sigma(\bar{y}_q - \bar{y})^2}{\Sigma(y - \bar{y})^2} \qquad (14.77)$$

Analogously to (14.75) we have

$$\Sigma(x - \beta_1 y)^2 = \Sigma(x - \bar{x}_p + \bar{x}_p - \beta_1 y)^2$$
$$= \Sigma(x - \bar{x}_p)^2 + \Sigma(\bar{x}_p - \beta_1 y)^2. \qquad (14.78)$$

But, from (14.11),

$$\Sigma(x - \beta_1 y)^2 = (1 - \rho^2)\Sigma(x - \bar{x})^2,$$

and from (14.76),

$$(1 - \eta^2_{xy})\Sigma(x - \bar{x})^2 = \Sigma(x - \bar{x}_p)^2.$$

Taking these results in conjunction with (14.78) we find

$$\Sigma(x - \bar{x})^2(\eta^2_{xy} - \rho^2) = \Sigma(\bar{x}_p - \beta_1 y)^2.$$

Hence η cannot be less than ρ. If and only if $\eta = \rho$, $\bar{x}_p - \beta_1 y$ vanishes for each array, i.e. the regression is linear. $\eta^2 - r^2$ may thus be used as an index of linearity of regression.

Example 14.11

The calculation of the correlation ratios is based on equation (14.76). As an illustration we will find those for the data of Table 14.1. The means of the horizontal arrays and the array frequencies are shown in Table 14.5.

TABLE 14.5

Calculation of the Correlation Ratio η_{xy} for the Data of Table 14.1.

Highest Audible Pitch	Frequency	Mean $\bar{x}_p$	$\bar{x}_p - \bar{x}$	$(\bar{x}_p - \bar{x})^2$
5–	3	4·666,667	3·896,025	15·179,011
7–	45	9·111,111	8·340,469	69·563,423
9–	10	9·700,000	8·929,358	79·733,434
11–	104	8·817,308	8·046,666	64·748,834
13–	93	6·333,333	5·562,691	30·943,531
15–	310	3·022,581	2·251,939	5·071,229
17–	576	1·064,236	0·293,594	0·086,197
19–	1051	0·101,808	− 0·668,834	0·447,339
21–	957	− 0·801,463	− 1·572,105	2·471,514
23–	165	− 1·278,788	− 2·049,430	4·200,163
25–	41	− 1·512,195	− 2·282,837	5·211,345
27–	16	− 1·562,500	− 2·333,142	5·443,552
29–	2	− 1·000,000	− 1·770,642	3·135,173
31–	2	− 3·000,000	− 3·770,642	14·217,741
33–	4	− 1·750,000	− 2·520,642	6·353,636

We have already found that

$$\Sigma(x^2) = 47{,}392 \qquad \Sigma(x) = 2604,$$

from which

$$\Sigma(x - \bar{x})^2 = \Sigma(x^2) - \frac{1}{N}(\Sigma(x))^2$$

$$= 45{,}385{\cdot}25.$$

From the table we now have

$$\Sigma(\bar{x}_p - \bar{x})^2 = 19{,}095{\cdot}88.$$

It should be noticed that in forming this sum we multiply each $(\bar{x}_p - \bar{x})^2$ in the last column of Table 14.5 by the corresponding frequency in the second column, for the summation takes place over all values of x.

We then find

$$\eta^2_{xy} = \frac{19{,}095{\cdot}88}{45{,}385{\cdot}25} = 0{\cdot}420{,}751,$$

giving $\eta_{xy} = 0{\cdot}6487$. Similarly it may be shown that $\eta_{yx} = 0{\cdot}6231$. The correlation coefficient is $-0{\cdot}6136$.

We have

$$\eta^2_{xy} - r^2 = 0{\cdot}044$$
$$\eta^2_{yx} - r^2 = 0{\cdot}012.$$

These values are close to zero and the regressions are thus approximately linear.

14.24. We shall see in the next chapter that η^2 is closely related to a statistic R, the multiple correlation coefficient, which is of rather greater importance. We accordingly defer a full discussion of the sampling distribution of η^2 until that chapter, but will here derive it in the special case of samples from an *uncorrelated* bivariate population.

From (14.75) and (14.76) we have

$$\frac{1 - \eta^2}{\eta^2} = \frac{\Sigma(x - \bar{x}_p)^2}{\Sigma(\bar{x}_p - \bar{x})^2} \quad . \quad . \quad . \quad . \quad (14.79)$$

Now if the population is normal and the arrays are of narrow width, the distribution in each array will be normal. We have already seen that in a normal distribution the mean is distributed independently of the variance. Hence $\Sigma(x - \bar{x}_p)^2$, which is the sum of numerical multiples of array variances, is independent of the array means and hence of the quantity $\Sigma(\bar{x}_p - \bar{x})^2$. Thus the numerator and denominator of (14.79) are independent.

Further, if the variates are uncorrelated and therefore (in the normal case) independent, the distributions in parent arrays have all the same mean and variance, those of the total distribution. Without loss of generality we may take the mean to be zero and the variance to be unity.

It was seen in Example 10.5 that the sum of squares of α variates, each distributed normally with zero mean and unit variance, is given by

$$dF \propto e^{-\frac{1}{2}t}\, t^{\frac{1}{2}(\alpha-2)}\, dt \quad . \quad . \quad . \quad . \quad . \quad (14.80)$$

and that the distribution of sum of squares about the mean is the same in form but has the index of t reduced by unity. Now $\Sigma(x - \bar{x}_p)^2$ summed over any given array containing N_p members is the sum of squares about the mean of N_p variates and is thus distributed in the form (14.80) with $N_p - 1$ degrees of freedom, that is to say with $\alpha = N_p - 1$.

Thus the sum of $(x - \bar{x}_p)^2$ for the whole array will be distributed in the form (14.80) with $\sum_p (N_p - 1) = N - p$ degrees of freedom, i.e. as

$$dF \propto e^{-\frac{1}{2}t} t^{\frac{1}{2}(N-p-2)} dt. \qquad (14.81)$$

The mean $\bar{x}_p$ will be distributed in the normal form

$$dF \propto e^{-\frac{1}{2}N_p \bar{x}_p^2} d\bar{x}_p$$

and consequently $\sum_x (\bar{x}_p - \bar{x})^2$, which is equal to $\sum_p N_p(\bar{x}_p - \bar{x})^2$ (the summation now extending over the p arrays), will be distributed in the form (14.80) with $p - 1$ degrees of freedom; i.e., writing u for the sum, as

$$dF \propto e^{-\frac{1}{2}u} u^{\frac{1}{2}(p-3)} du. \qquad (14.82)$$

To find the distribution of $\frac{1 - \eta^2}{\eta^2}$ we then have to find that of $\frac{t}{u}$, t and u being independent.

We have for the joint distribution

$$dF \propto \exp\left[-\tfrac{1}{2}(t + u)\right] t^{\frac{1}{2}(N-p-2)} u^{\frac{1}{2}(p-3)} dt\, du \qquad (14.83)$$

Put $\xi = \frac{t}{u}$ $\qquad \zeta = t + u.$

The Jacobian of the transformation is

$$\frac{\partial(\xi, \zeta)}{\partial(t, u)} = \frac{t + u}{u^2}$$

and (14.83) becomes

$$dF \propto e^{-\frac{1}{2}\zeta} \zeta^{\frac{1}{2}(N-3)} d\zeta \frac{\xi^{\frac{1}{2}(N-p-2)}}{(1 + \xi)^{\frac{1}{2}(N-1)}} d\xi.$$

Thus ξ and ζ are independent and we have for the distribution of ξ

$$dF \propto \frac{\xi^{\frac{1}{2}(N-p-2)}}{(1 + \xi)^{\frac{1}{2}(N-1)}} d\xi \qquad (14.84)$$

whence, on putting $\xi = \frac{1 - \eta^2}{\eta^2}$ we find

$$dF \propto (1 - \eta^2)^{\frac{1}{2}(N-p-2)} (\eta^2)^{\frac{p-3}{2}} d(\eta^2)$$

$$= \frac{1}{B\left(\frac{N-p}{2}, \frac{p-1}{2}\right)} (1 - \eta^2)^{\frac{1}{2}(N-p-2)} (\eta^2)^{\frac{1}{2}(p-3)} d(\eta^2) \qquad (14.85)$$

which is the distribution required.

14.25. The distribution function of (14.85), which is a Pearson Type I curve, may be obtained from the incomplete B-function. It is sufficient for ordinary purposes, however, to use the tabulated forms of Fisher's z-distribution (Example 10.18). In fact, putting in (14.85)

$$\nu_1 = p - 1$$
$$\nu_2 = N - p$$
$$e^{2z} = \frac{\eta^2}{1 - \eta^2} \frac{N - p}{p - 1}$$

we find

$$dF \propto \frac{e^{\nu_1 z}}{(\nu_1 e^{2z} + \nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}},$$

the form of equation (10.62). Appendix Tables 4 and 5 give the values of z, such that equal or greater values will be attained with probability 0·05 and 0·01. These tables are due to Fisher and reproduced from his *Statistical Methods for Research Workers*. In practice, however, η^2 is only calculated for large values of N outside the range of these tables, and we may either use the approximation suggested therein or special Tables by T. L. Woo reproduced in *Tables for Statisticians and Biometricians*, Part II.

14.26. It is easy to show that the first two moments of (14.85) and the constants γ_1 and γ_2 are given by

$$\mu_1' = \frac{p-1}{N-1} \tag{14.86}$$

$$\mu_2 = \frac{2(p-1)(N-p)}{(N-1)^2(N+1)} \tag{14.87}$$

$$\gamma_1^2 = \frac{16(N-2p+1)^2(N+1)}{(p-1)(N-p)(N+3)^2} \tag{14.88}$$

$$\gamma_2 = \frac{6(N+1)\{2(N-1)^2 + (p+1)(N-p)(N-13)\}}{(p-1)(N-p)(N+3)(N+5)} \tag{14.89}$$

Thus, to order N^{-1},

$$\left.\begin{aligned} \gamma_1^2 &\sim \frac{16}{p-1} \\ &\quad\ \frac{12}{p-1} \end{aligned}\right\} \tag{14.90}$$

and thus η^2 does not tend to normality for large N for any finite number of arrays p.

Tetrachoric r

14.27. We now proceed to consider two coefficients designed for the measurement of dependence and based on the product-moment correlation coefficient, tetrachoric r and biserial η. Both those coefficients are, in effect, estimates of a putative product-moment correlation for data which are not specified with the detail of an ordinary bivariate table.

Suppose we have a fourfold table

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

. (14.91)

If this table is derived by a double dichotomy of a bivariate frequency-distribution

$$z = z_0 \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right.\right.$$

we may ask, what is the value of ρ in terms of a, b, c, d and N? This problem is, in fact, determinate.

If the population is normal the array totals will be normal, and thus the frequencies $(a+c)$ and $(b+d)$ correspond to a dichotomy of the normal curve, i.e. there exists an h' such that

$$\left.\begin{aligned}\int_{-\infty}^{h'}\int_{-\infty}^{\infty} z\,dx\,dy &= \frac{a+c}{N}\\ \int_{h'}^{\infty}\int_{-\infty}^{\infty} z\,dx\,dy &= \frac{b+d}{N}\end{aligned}\right\} \tag{14.92}$$

or

$$\frac{1}{\sigma_1\sqrt{(2\pi)}}\int_{-\infty}^{h'} \exp\left\{-\tfrac{1}{2}\frac{x^2}{\sigma_1^2}\right\}dx = \frac{a+c}{N}$$

$$\frac{1}{\sigma_1\sqrt{(2\pi)}}\int_{h'}^{\infty} \exp\left\{-\tfrac{1}{2}\frac{x^2}{\sigma_1^2}\right\}dx = \frac{b+d}{N}.$$

Putting $h = \dfrac{h'}{\sigma_1}$ we have

$$\left.\begin{aligned}\int_{-\infty}^{h} \exp(-\tfrac{1}{2}x^2)\,dx &= \frac{(a+c)\sqrt{(2\pi)}}{N}\\ \int_{h}^{\infty} \exp(-\tfrac{1}{2}x^2)\,dx &= \frac{(b+d)\sqrt{(2\pi)}}{N}\end{aligned}\right\} \tag{14.93}$$

so that h can be derived from the tables of the normal integral.

Similarly there will be a k defined by

$$\int_{-\infty}^{k} \exp(-\tfrac{1}{2}y^2)\,dy = \frac{(a+b)\sqrt{(2\pi)}}{N}.$$

We then require to solve for ρ the equation

$$\frac{d}{N} = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\int_h^{\infty}\int_k^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]dx\,dy \quad . \tag{14.94}$$

We will expand the integral on the right in ascending powers of ρ. The characteristic function of the distribution is

$$\phi(t, u) = \exp\{-\tfrac{1}{2}(t^2 + 2\rho tu + u^2)\}.$$

Thus

$$\frac{d}{N} = \frac{1}{4\pi^2}\int_h^{\infty} dx\int_k^{\infty} dy\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(t, u)e^{-itx-iuy}\,dt\,du$$

$$= \frac{1}{4\pi^2}\int_h^{\infty} dx\int_k^{\infty} dy\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(t^2+u^2) - itx - iuy\}\sum_{j=0}^{\infty}\frac{(-\rho)^j t^j u^j}{j!}\,dt\,du \tag{14.95}$$

The coefficient of $(-\rho)^j$ is the product of two integrals, the first of which is

$$\frac{1}{2\pi}\int_h^{\infty} dx\int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}t^2 - itx)t^j\,dt$$

and the second a similar expression in k, y and u. Now from **6.24** the integral with respect to t is equal to

$$(-i)^j H_j(x) . \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

and hence the double integral is

$$\left[(-1)^{j-1}\, i^j\, H_{j-1}(x)\, \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right]_h^\infty .$$

Hence, from (14.95),

$$\frac{d}{N} = \Sigma \frac{\rho^j}{j!} H_{j-1}(h) . H_{j-1}(k) . \frac{1}{2\pi} e^{-\frac{1}{2}h^2} e^{-\frac{1}{2}k^2}.$$

In the notation of **6.27** we write τ for the tetrachoric function of h and τ' for that of k, and we then have

$$\frac{l}{N} = \sum_{j=0}^{\infty} \rho^j \tau_j \tau_j \qquad (14.96)$$

The tetrachoric functions have been tabled up to τ_{19} (*Tables for Statisticians and Biometricians*, Parts I and II) and, with their aid, (14.96) can be solved by successive approximation. Examples will be found in the introduction to the Tables.

14.28. It is to be realised that the coefficient obtained by the solution of equation (14.96) is not a product–moment correlation, but an estimate of the parameter ρ in a bivariate normal population. It is not an estimate of the product–moment correlation in non-normal populations. Its practical use is limited largely by arithmetical inconvenience, both in the solution of (14.96) and in the determination of sampling variances. Karl Pearson (1913) has given expressions for these quantities, but as nothing is known of the distribution of tetrachoric r it is not clear how far the use of a standard error is justifiable.

Biserial η

14.29. Suppose now that we have a $2 \times q$-fold table, the dichotomy being according to some qualitative factor and the other classification either to a numerical variate or to some variate permitting the arrangement of the classes in order.

Table 14.6 will illustrate the type of material under discussion. The data relate to

TABLE 14.6

Showing 1426 Criminals classified according to Alcoholism and Type of Crime.

(C. Goring's data, quoted by K. Pearson, 1909.)

	Arson.	Rape.	Violence.	Stealing.	Coining.	Fraud.	Totals.
Alcoholic	50	88	155	379	18	63	753
Non-alcoholic . . .	43	62	110	300	14	144	673
Totals	93	150	265	679	32	207	1426

1426 criminals classified according to whether they were alcoholic or not and according to the crime for which they were imprisoned. The order of the crime-classification is deter-

mined by its relationship with intelligence, arson being associated with low intelligence and fraud with high.

If the population is normal, $\eta = \rho$. We have

$$\eta^2_{yx} = \frac{\Sigma N_p (\bar{y}_p - \bar{y})^2}{N \text{ var } y}$$

$$= \Sigma \frac{(N_p \bar{y}_p^2 - 2N_p \bar{y}_p \bar{y} + N_p \bar{y}^2)}{N \text{ var } y}$$

$$= \Sigma\left(\frac{N_p \bar{y}_p^2}{N \text{ var } y}\right) \quad \frac{\bar{y}^2}{\text{var } y} \qquad . \quad (14.97)$$

Since

$$\Sigma\left(\frac{N_p \bar{y}_p \bar{y}}{N \text{ var } y}\right) = \frac{\bar{y}}{N \text{ var } y} \qquad \Sigma N_p \bar{y}_p = \frac{\bar{y}^2}{\text{var } y}.$$

Thus

$$\eta^2 = \frac{1}{N} \Sigma \frac{N_p y_p^2}{\text{var } y_p} \cdot \frac{\text{var } y_p}{\text{var } y} - \frac{y^2}{\text{var } y} \qquad . \quad . \quad . \quad . \quad (14.98)$$

But the mean variance of arrays, weighted according to the numbers in arrays, $= \text{var } y(1 - \rho^2) = \text{var } y(1 - \eta^2)$. Taking this as equal to var y_p we have

$$\eta^2 = (1 - \eta^2) \Sigma\left(\frac{N_p \bar{y}_p^2}{N \text{ var } y_p}\right) - \frac{\bar{y}^2}{\text{var } y}$$

giving

$$\eta^2 = \frac{\frac{1}{N} \Sigma\left(\frac{N_p \bar{y}_p^2}{\text{var } y_p}\right) - \frac{\bar{y}^2}{\text{var } y}}{1 + \frac{1}{N} \Sigma\left(\frac{N_p \bar{y}_p^2}{\text{var } y_p}\right)} \qquad . \qquad (14.99)$$

The use of this expression lies in the fact that the quantities in it can be estimated from the data on certain assumptions. If we suppose that the quantity according to which dichotomy has been made (in our example, alcoholism) is capable of representation by a variate which is normally distributed, and thus that each y-array is a dichotomy of a normal curve, the quantities $\frac{\bar{y}}{\sqrt{\text{var } y}}$ and $\frac{\bar{y}_p}{\sqrt{\text{var } y_p}}$ can be obtained from the tables of the normal integral. For example, the two frequencies alcoholic and non-alcoholic are, for arson, 50 and 43. Thus the proportional frequency in the alcoholic group is $\frac{50}{93} = 0{\cdot}5376$ and the deviation corresponding to this frequency is seen from the tables to be 0·0946, which is thus $\frac{\bar{y}_p}{\sqrt{\text{var } y_p}}$ for this y-array.

Example 14.12

For the data of Table 14.6 the proportional frequencies, the values of $\frac{\bar{y}_p}{\sqrt{\text{var}}}$ and the N_p are as follows :—

	Arson.	Rape.	Violence.	Stealing.	Coining.	Fraud.	TOTAL.
Alcoholic .	0·5376	0·5867	0·5849	0·5582	0·5625	0·3043	0·5281
$\bar{y}_p / \sqrt{\text{var } y_p}$	0·0944	0·2190	0·2144	0·1463	0·1573	− 0·5119	0·0704
N_p . .	93	150	265	679	32	207	1426

Then from (14.99) we have

$$\eta^2 = \frac{\frac{1}{1426}\{93(0{\cdot}0944)^2 + \ldots\} - (0{\cdot}0704)^2}{1 + \frac{1}{1426}\{93(0{\cdot}0944)^2 + \ldots\}}$$

giving

$$\eta^2 = 0{\cdot}05456$$
$$\eta = 0{\cdot}234,$$

which, on our various assumptions, may be taken as approximating to the supposed product-moment correlation coefficient.

As for tetrachoric r, the sampling distribution of biserial η is unknown. Expressions for its sampling variance have been derived by K. Pearson (1915), but are to be used with considerable reserve.

14.30. Something may also be said about the assumptions on which tetrachoric r and biserial η are based, particularly that of normality. In supposing that a given fourfold table is the double dichotomy of a normal population, we are assuming that the attributes or variates concerned are capable of representation on a normal scale and that it was, in fact, this scale which determined the classification given. This assumption is evidently a considerable one and cannot always be made with much confidence. In dividing criminals into alcoholic and non-alcoholic it would, for example, be assumed that "alcoholism" is a quantity which varies continuously from one subject to another; or perhaps that propensity to alcoholism was such a variate. At one end of the scale we should have chronic inebriety, at the other the most austere teetotalism. It would be further assumed that if the degree of alcoholism could be measured, the population of criminals would be distributed according to the alcoholic variate in a normal form; and it would be further assumed that the data which are given would have been arrived at by a dichotomy of the population according to the variate. How far assumptions of this kind are justified depends on previous knowledge and the circumstances of individual cases; but even so it remains largely a matter of personal opinion. The reader will meet widely divergent views in the literature of the subject.

Intra-class Correlation

14.31. There sometimes arise, mainly in biological work, cases in which we require the correlation between members of one or more families. We might, for example, wish to examine the correlation between heights of brothers. The question then arises, which is the first variate and which the second? In the simplest case we might have a number of families each containing two brothers. Our correlation table has two variates, both height, but in order to complete it we must decide which brother is to be related to which variate. One way of doing so would be to take the elder brother first, or the taller brother; but this would provide the answer to different questions, the correlation between elder and younger brothers, or between taller and shorter brothers; not the correlation between brothers in general.

The problem is met by entering in the correlation table both possible pairs, i.e. those obtained by taking both brothers first. Generally, if the family contains k members, there will be $k(k-1)$ entries, each member being taken first in association with each other

member second. If there are p families with $k_1, k_2 \ldots k_p$ members there will be $\sum_{i=1}^{p} k_i(k_i - 1)$ entries in the correlation table. As a simple illustration consider five families of three brothers with heights

1st family	69, 70, 72 inches
2nd family	70, 71, 72 ,,
3rd family	71, 72, 72 ,,
4th family	68, 70, 70 ,,
5th family	71, 72, 73 ,,

There will be 30 entries in the table, which will be as follows :—

Height (inches).		Height (inches).						
		68	69	70	71	72	73	TOTALS.
	68	—	—	2	—	—	—	2
	69	—	—	1	—	1	—	2
	70	2	1	2	1	2	—	8
	71	—	—	1	—	4	1	6
	72	—	1	2	4	2	1	10
	73	—	—	—	1	1	—	2
	TOTALS	2	2	8	6	10	2	30

Here, for example, the pair 69, 70 in the first family is entered as (69, 70) and (70, 69) and the pair 72, 72 in the third family *twice* as (72, 72).

The table is symmetrical about the diagonal, as it evidently must be. We may calculate the product-moment coefficient in the usual way. We find var x = var y = 1·716, cov (xy) = 0·516 and hence $\rho = \dfrac{0{\cdot}516}{1{\cdot}716} = 0{\cdot}301$.

The actual compilation of such a table is, however, both tedious and unnecessary. The coefficient ρ can be found by direct methods, as follows :—

Suppose there are p families, with variate-values $x_{11} \ldots x_{1k_1}, x_{21} \ldots x_{2k_2}, \ldots x_{p1} \ldots x_{pk_p}$, the families numbering $k_1, k_2 \ldots k_p$. In the correlation table each member of the ith family will appear $k_i - 1$ times (in association with the other members of the family), and thus the mean of each variate is given by

$$\bar{x} = \bar{y} = \frac{1}{N} \sum_i \{(k_i - 1) \sum_j (x_{ij})\}. \tag{14.100}$$

the first summation taking place over the p families and the second over all members of the ith family. Similarly

$$\operatorname{var} x = \operatorname{var} y = \frac{1}{N}\sum_i\{(k_i - 1)\sum_j'(x_{ij} - \bar{x})^2\} \quad . \quad . \quad . \quad (14.101)$$

and

$$\operatorname{cov}(xy) = \frac{1}{N}\sum_i\sum_{j,l}(x_{ij} - \bar{x})(x_{il} - \bar{x}), \qquad j \neq l, \quad . \quad . \quad . \quad (14.102)$$

the summation $\sum\limits_{j,l}$ extending over all possible pairs for which $j \neq l$. Thus the coefficient ρ is given by

$$\rho = \frac{\sum\limits_i\sum\limits_{j,l}(x_{ij} - \bar{x})(x_{il} - \bar{x})}{\sum\limits_i(k_i - 1)\sum\limits_j'(n_{ij} - \bar{x})^2} \quad . \quad . \quad . \quad . \quad . \quad (14.103)$$

This can be thrown into a rather more convenient form. We have

$$\sum_i\sum_{j,l}(x_{ij} - \bar{x})(x_{il} - \bar{x}) = \sum_i\sum_{il}'(x_{ij} - \bar{x})(x_{il} - \bar{x}) - \sum_i\sum_j(x_{ij} - \bar{x})^2$$

(where the summation $\sum\limits_{il}'$ now extends over all possible pairs, including $j = l$)

$$= \sum_i k_i^2(\bar{x}_i - \bar{x})^2 - \sum_i\sum_j(x_{ij} - \bar{x})^2, \qquad (14.104)$$

$\bar{x}_i$ being the mean of the ith family.
Thus

$$= \frac{\sum\limits_i k_i^2(\bar{x}_i - \bar{x})^2 - \sum\limits_i\sum\limits_j(x_{ij} - \bar{x})}{\sum(k_i - 1)\sum(x_{ij} - \bar{x})^2} \qquad (14.105)$$

If all the families have the same number of members this formula is somewhat simplified. Denoting by v the variance of x, and by v_m the variance of means of families (about the mean $\bar{x}$), we have

$$\rho = \frac{pk^2v_m - pkv}{(k - 1)pkv}$$

$$= \frac{1}{(k - 1)}\left(\frac{kv_m}{v} - 1\right) \quad . \quad . \quad . \quad . \quad . \quad (14.106)$$

The coefficient ρ is called the Intra-class Correlation Coefficient, to distinguish it from the ordinary product-moment coefficient.

Example 14.13

Let us use formula (14.106) to find the intra-class coefficient for the example of the previous section. With a working mean at 70 inches, the values of the variates are

$$-1, 0, 2; \quad 0, 1, 2; \quad 1, 2, 2; \quad -2, 0, 0; \quad 1, 2, 3.$$

Hence $\bar{x} = \frac{13}{15}$. $\mu_2' = \frac{1}{15}\{(-1)^2 + 0^2 + \ldots\} = \frac{37}{15}$

$$\operatorname{var}(x) = \frac{386}{225}.$$

The means of families are

$$\frac{5}{15}, \quad \frac{15}{15}, \quad \frac{25}{15}, \quad \frac{-10}{15}, \quad \frac{30}{15},$$

and the deviations from $\bar{x}$

$$\frac{-8}{15},\quad \frac{2}{15},\quad \frac{12}{15},\quad \frac{-23}{15},\quad \frac{17}{15}$$

Thus

$$v_m = \frac{1}{5}\left\{\left(\frac{8}{15}\right)^2 + \text{etc.}\right\}$$
$$= \frac{1030}{1125}.$$

Hence, from (14.106)

$$\rho = \frac{1}{2}\left\{\frac{3.1030.225}{1125.386} - 1\right\}$$
$$= 0{\cdot}301,$$

a result we have already found directly.

14.32. One caution is necessary in the interpretation of the intra-class correlation coefficient. From (14.106) it is seen that intra-class ρ cannot be less than $\frac{1}{k-1}$, though it may attain $+1$. It is thus a skew coefficient in the sense that, unlike product–moment correlation and association, a negative value has not the same significance (as a departure from independence) as the equivalent positive value.

14.33. The sampling distribution of intra-class ρ for the case of a normal population and equal numbers in families may be obtained as follows:—

It may be shown, precisely as in (14.25), that the ratio of two sums of squares about means, $\xi = \frac{v_1}{v_2}$, based on $N - p$ and $p - 1$ sums, is distributed as

$$dF \propto \frac{\xi^{\frac{1}{2}(N-p-2)}\, d\xi}{\left(1 + \frac{\sigma_2^2 \xi}{\sigma_1^2}\right)^{\frac{1}{2}(N-1)}}, \qquad (14.107)$$

provided that the sums are independent and emanate from normal populations. Here σ_2^2, σ_1^2 are the population variances relating to v_2, v_1 respectively.

Consider now p families of k members, pk in all, as p samples of k from a normal population in which the intra-class coefficient is λ. Writing l for the sample intra-class coefficient we have

$$l = \frac{1}{k-1}\left(\frac{kv_m}{v} - 1\right)$$
$$= \frac{1}{k-1}\left(\frac{k}{1+\xi} - 1\right) \qquad (14.108)$$

where $\xi = \frac{v - v_m}{v_m}$. Now v_m relates to means of samples and is distributed independently of $v - v_m$, as in the case of (14.25). We may therefore substitute for ξ in (14.107), with $N = pk$ and $p = p$. Furthermore, since the population value of $v - v_m$ is σ_1^2 and that of v_m is $\frac{\sigma_2^2}{k-1}$, we have

$$\lambda = \frac{1}{k-1}\left\{\frac{k\sigma_2^2}{(k-1)\,\sigma_1^2 + \sigma_2^2} - 1\right\}$$
$$= \frac{\sigma_2^2 - \sigma_1^2}{(k-1)\,\sigma_1^2 + \sigma_2^2} \qquad (14.109)$$

After a little reduction (14.107) becomes

$$dF \propto \frac{(1-l)^{\frac{p(k-1)-2}{2}}\{1+l(k-1)\}^{\frac{p-3}{2}}\,dl}{\{1-\lambda+\lambda(k-1)(1-l)\}^{\frac{kp-1}{2}}} \quad . \quad . \quad . \quad . \quad (14.110)$$

As for the product–moment coefficient, this form may be brought closer to normality by putting

$$l = \tanh z, \quad \lambda = \tanh \zeta.$$

In the particular case $k = 2$ we find

$$\begin{aligned} dF &\propto \frac{e^{-\frac{1}{2}z}\operatorname{sech}^{p-\frac{1}{2}} z\, dz}{\cosh^{p-\frac{1}{2}}(z-\zeta)} \\ &= \frac{\Gamma(n-\frac{1}{2})}{\Gamma(n-1)\sqrt{(2\pi)}} \operatorname{sech}^{n-\frac{1}{2}}(z-\zeta)e^{-\frac{1}{2}(z-\zeta)} \quad . \quad . \quad . \quad . \quad (14.111) \end{aligned}$$

which has the remarkable property of depending only on $z - \zeta$, i.e. of being the same in form for any ζ or λ. Writing $z - \zeta = x$ we may derive from (14.111) the expansion

$$y = \frac{\Gamma(n-\frac{1}{2})}{\Gamma(n-1)\sqrt{(2\pi)}} e^{-\frac{n-1}{2}x^2}\left[1 + \frac{n-1}{12}x^4 \right.$$
$$\left. \frac{n-1}{45}x^6 - \frac{(n-1)^2}{288}x^8\Big\}\right]\left[1 - \tfrac{1}{2}x - \frac{x^2}{8} + \frac{5x^3}{48} + \frac{17x^4}{384} \ldots\right]. \quad (14.112)$$

giving

$$\mu_1' = -\frac{1}{2(n-1)}\Big\{1 + \frac{1}{2(n-1)} + \quad (14.113)$$

$$\mu_2 = \frac{}{n-1}\Big\{1 + \frac{1}{2(n-1)} + \frac{1}{6(n-1)^2} + \ldots \quad (14.114)$$

$$\mu_3 = -\frac{1}{(n-1)^3} + \ldots \quad (14.115)$$

$$\mu_4 = \frac{1}{(n-1)^2}\Big\{3 + \frac{5}{n-1} + \frac{19}{4(n-1)^2} \ldots \quad (14.116)$$

whence

$$\gamma_1 = \frac{}{(n-1)^{\frac{3}{2}}} \quad (14.117)$$

$$\gamma_2 = \frac{2}{n-1} + \frac{}{(n-1)^2} \quad . \quad . \quad . \quad . \quad (14.118)$$

illustrating the tendency to normality. $z - \zeta$ may be taken to be distributed normally about zero mean with variance $\dfrac{1}{(n-\frac{3}{2})}$ approximately.

For the general case the substitution

$$\left.\begin{aligned} 2(k-1)l &= k - 2 + k\tanh(z-\theta) \\ 2(k-1)\lambda &= k - 2 + k\tanh(\zeta-\theta) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (14.119)$$

where $\tanh\theta = \dfrac{k-2}{k}$, reduces (14.110) to

$$\frac{\Gamma\left(\frac{kp-1}{2}\right)}{2^{\frac{kp-3}{2}}\Gamma\{(k-1)p-2\}\Gamma\left(\frac{n-1}{2}\right)} \exp\left\{-\frac{(k-2)p+1}{}(x-\theta)\right.$$
$$\times \operatorname{sech}\frac{pk-1}{2}(x-\theta)\,dx \quad . \quad . \quad (14.120)$$

where, as usual, $x = z - \zeta$

NOTES AND REFERENCES

The classical theory of product–moment correlation, beginning with Galton and Karl Pearson, was established by Yule (1897, 1907). The sampling problem for the normal case was solved by Fisher (1915) and studied by subsequent writers, culminating in Miss David's tables of 1938. For experimental work on the sampling distribution see E. S. Pearson (1931). A method of deriving the distribution alternative to the geometrical approach and relying on characteristic functions has been given by Kullback (1935).

Tetrachoric ρ and biserial η are both inventions of Karl Pearson's, but the tetrachoric series has been discovered by many writers, priority apparently being due to Mehler (1876). For controversy on the nature and scope of tetrachoric ρ see references in previous chapter.

Intra-class correlation is formally equivalent to a linear function of the ratio of two variances and thus becomes a branch of quadratic analysis (analysis of variance) which will be dealt with in the second volume.

Co-operative Study (1917) (H. E. Soper, A. W. Young, B. M. Cave, A. Lee and K. Pearson), "On the distribution of the correlation coefficient in small samples," *Biometrika*, **11**, 328.

David, F. N. (1938), *Tables of the Correlation Coefficient*, Cambridge University Press.

Fisher, R. A. (1915), "The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, **10**, 507.

—— (1921), "On the probable error of a coefficient of correlation deduced from a small sample," *Metron*, **1**, No. 4, 3.

Kullback, S. (1934), "An application of characteristic functions to the distribution problem of statistics," *Ann. Math. Statist.*, **5**, 263.

Mehler, G. (1876), "Reihenentwicklung nach Laplaceschen Functionen höherer Ordnung," *J. für Math.*, **66**, 161.

Pearson, K. (1909), "On a new method of determining correlation between a measured characteristic A and a character B, etc.," *Biometrika*, **7**, 96.

—— (1910), "On a new method of determining correlation when one variable is given by alternative and the other by multiple categories," *Biometrika*, **7**, 248.

—— (1911), "On the correction necessary for the correlation ratio η," *Biometrika*, **8**, 254, and (1923), **14**, 412.

—— (1913), "On the probable error of a coefficient of correlation as found from a fourfold table," *Biometrika*, **9**, 22.

—— (1915), "On the probable error of biserial η," *Biometrika*, **11**, 292.

—— and Pearson, E. S. (1922), "On polychoric coefficients of correlation," *Biometrika*, **14**, 127.

Pearson, E. S. (1931), "The test of significance for the correlation coefficient," *Jour. Amer. Statist. Ass.*, **26**, 128, and (1932) **27**, 424. See also Cheshire, L., Oldis, E., and Pearson, E. S. (1932), *ibid.*, **27**, 121.

Ritchie-Scott, A. (1918), "The correlation coefficient of a polychoric table," *Biometrika*, **12**, 93.

Yule, G. U. (1897), "On the theory of correlation," *Jour. Roy. Statist. Soc.*, **60**, 812.

—— (1907), "On the theory of correlation for any number of variables treated by a new system of notation," *Proc. Roy. Soc.*, A, **79**, 182.

EXERCISES

14.1. Show that the data of Table 1.25 have the following constants (x = age, y = milk yield):

$$\text{mean } x = 6{\cdot}22 \text{ years.} \qquad \text{mean } y = 18{\cdot}61 \text{ gallons.}$$
$$\sqrt{(\text{var } x)} = 2{\cdot}21 \quad ,, \qquad \sqrt{(\text{var } y)} = 3{\cdot}37 \quad ,,$$
$$\rho = 0{\cdot}219, \quad \eta_{xy} = 0{\cdot}242, \quad \eta_{yx} = 0{\cdot}266.$$

14.2. Show that for the data of Table 14.2

$$\rho = -\,0{\cdot}014, \quad \eta_{xy} = 0{\cdot}14, \quad \eta_{yx} = 0{\cdot}38.$$

14.3. Show that the smaller angle between the regression lines is

$$\text{arc tan } \frac{1}{\rho} \; \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}.$$

14.4. If a bivariate normal surface is dichotomised at its medians and α is the proportional frequency in the positive compartment of the 2×2 table so generated (i.e. the compartment including the limits $+\infty$), show that

$$\rho = \cos (1 - 2\alpha)\pi.$$

(Sheppard, *Phil. Trans. Roy. Soc.*, 1898, **192A**, 101.)

14.5. Show that the ordinates of the sampling distribution of the correlation coefficient r in samples from a normal parent with correlation ρ obey the recurrence relation

$$y_{n+2} = \frac{2n-1}{n-1}\alpha_1 y_{n+1} + \frac{n-1}{n-2}\alpha_2 y_n,$$

where n is the sample number and

$$\alpha_1 = \frac{\rho r \sqrt{(1-\rho^2)}\sqrt{(1-r^2)}}{\rho^2 r^2} \qquad \alpha_2 = \frac{(1-\rho^2)(1-r^2)}{1-\rho^2 r^2}.$$

(Co-operative Study, 1917.)

14.6. By the transformation $\cosh z - \rho r = \dfrac{1 - \rho r}{1 + u}$ show that the ordinate of the distribution of r may be expressed as

$$y_n = \frac{n-2}{\sqrt{(2\pi)}} \frac{(1-\rho^2)^{\frac{n-1}{2}} (1-r^2)^{\frac{n-4}{2}}}{(1-\rho r)^{\frac{2n-3}{2}}} \frac{\Gamma(n-1)}{\Gamma(n-\frac{1}{2})}$$
$$\left\{ \frac{1^2}{2^2}\frac{\alpha}{n-\frac{1}{2}} + \frac{1^2.3^2}{2^2.2^2}\frac{\alpha^2}{(n-\frac{1}{2})(n+\frac{1}{2})} + \frac{1^2.3^2.5^2}{2^2.2^2.2^2}\frac{\alpha^3}{(n-\frac{1}{2})(n+\frac{1}{2})(n-\frac{1}{2})} + \ldots \right\}$$

where $\alpha = \dfrac{1 + \rho r}{[\text{illegible}]}$.

14.7. Show that the characteristic function of

$$\theta_1 = \frac{\Sigma x^2}{2(1-\rho^2)\sigma_1^2}, \quad \theta_2 = -\frac{\rho \Sigma xy}{(1-\rho^2)\sigma_1\sigma_2}, \quad \theta_3 = \frac{\Sigma(y^2)}{2(1-\rho^2)\sigma_2^2}$$

in normal samples is

$$\frac{(1-\rho^2)^{\frac{n}{2}}}{\{(1-it_1)(1-it_3)-\rho^2(1+it_2)^2\}^{n}},$$

where t_1 refers to θ_1, and so on. Hence show that the distribution of variances and co-variances has the same characteristic function, except for constants, but with the value of n reduced by unity. Show that the simultaneous distribution of these quantities is then that of equation (14.42) with $\theta_1 = na$, $\theta_2 = n\rho b$, $\theta_3 = nc$.

(Kullback, 1934.)

14.8. From the distribution of equation (14.42) show that the distribution of $v = \frac{s_1}{\sigma_1}\Big/\frac{s_2}{\sigma_2}$ and r is given by

$$dF \propto \frac{v^{n-2}(1-r^2)^{\frac{n-4}{2}}}{(1-2\rho r v+v^2)^{n-1}}\,dr\,dv.$$

Integrating for r from -1 to $+1$ by putting

$$r = \frac{u(\lambda+\mu)-(1-u)(\lambda-\mu)}{u(\lambda+\mu)+(1-u)(\lambda-\mu)},\quad \lambda = 1+v^2,\quad \mu = 2\rho v,$$

show that the distribution of v is

$$dF = \frac{2(1-\rho^2 [illegible] ^{n-1} [illegible] ^{n-2}}{B([illegible] 1)\,(1+v^2)^{n-1}}\left\{1-\frac{4\rho^2 v^2}{(1+v^2)^2}\right\}^{-\frac{n}{2}}.$$

(This gives the distribution of the variance ratio when the variates are correlated. The result is due to S. S. Bose (1935), *Sankhyā*, **2**, 65. The derivation was given by Finney, *Biometrika*, 1938, **30**, 190.)

14.9. Show that in samples from a normal bivariate population the variance of b_2 is given exactly by

$$\mathrm{var}\,(b_2) = \frac{[illegible]}{n-3}\,\frac{\sigma_2^2}{\sigma_1^2}(1-\rho^2)$$

and that for the distribution of b_2

$$\gamma_1 = 0$$

$$\gamma_2 = \frac{[illegible]}{n-5}$$

14.10. By considering the joint distribution of s_1 and b_2 in normal samples, show that the regression of b_2 on s_1 is linear, but that of s_1 on b_2 is not linear and does not tend to linearity for large samples.

14.11. Writing the bivariate frequency function in the form

$$f(x, y) = f(x)\; g_x(y),$$

so that the jth moment about the origin of the y array for given x is

$$\mu_j'(x) = \int_{-\infty}^{\infty} dy\; y^n\; g_x(y),$$

show that

$$\left[\frac{\partial^n \phi(t, u)}{\partial u^n}\right]_{u=0} = i^j \int_{-\infty}^{\infty} dx\, e^{-itx} f(x)\, \mu_j'(x)$$

(where ϕ is the characteristic function of the distribution) so that

$$f(x)\, \mu_j'(x) = \frac{(-i)^j}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \left[\frac{\partial^n \phi}{\partial u^n}\right]_{u=0} dt.$$

Verify that the bivariate normal distribution has linear regressions and is homoscedastic.

14.12. (Data of E. M. Elderton, quoted by K. Pearson, 1910.) The following table shows 811 sons classified according to alcoholism of parent and health of son :—

Son.

	Healthy.	Fairly healthy.	Delicate.	Phthisical or epileptic.	Died young.	TOTALS.
Alcoholic . . .	122	9	24	8	42	205
Non-alcoholic .	328	37	71	37	133	606
TOTALS	450	46	95	45	175	811

Show that biserial $\eta = 0{\cdot}089$, indicating little correlation between health of son and consumption of alcohol by parent.

14.13. (Data from O. H. Latter, *Biometrika*, **4**, 1905, p. 363.)
The following table shows the length of cuckoos' eggs fostered by various birds :—

Length of Egg (units $\frac{1}{2}$ millimetre).

Foster Parent.	40	41	42	43	44	45	46		48	49	50	TOTALS.
Robin	1	1	8	3	9	13	20		11			
Wren	7	5	14	8	9	6	3					54
Hedge-Sparrow . .	—	—		5	14	13	13		5	—	3	58
TOTALS	8	6	24	16	32	32	36	11	16	2		188

Show that the coefficient of intra-class correlation is $+\,0{\cdot}22$.

14.14. A series of measurements are subject to errors of observation which may be supposed uncorrelated with the magnitudes of the measurements. If x_1, y_1 refer to the

observed deviations from arithmetic means and x, y to the true deviations, show that $\Sigma(x_1y_1) = \Sigma(xy)$, but that var $x_1 >$ var x ; var $y_1 >$ var y. Hence show that the observed correlation is less than the true correlation.

14.15. If three variables X_1, X_2, X_3 are uncorrelated and the deviations are small compared with their mean values M_1, M_2 and M_3, show that the variance of $\frac{X_1}{X_2}$ is approximately

$$\frac{M_1^2}{M_2^2}\left(\frac{\text{var } X_1}{M_1^2} - \frac{2 \text{ cov } (X_1, X_2)}{M_1 M_2} + \frac{\text{var } X_2}{M_2^2}\right),$$

and that the correlation between $\frac{X_1}{X_3}$ and $\frac{X_2}{X_3}$ is

$$\rho = \frac{}{\sqrt{(v_1^2 + v_3^2)(v_2^2 + v_3^2)}},$$

where $v_1^2 = \frac{\text{var } X_1}{M_1{}^2}$, etc.

Note that this is positive, so that there is a "spurious" correlation between the two indices $\frac{X_1}{X_3}$ and $\frac{X_2}{X_3}$.

CHAPTER 15

PARTIAL AND MULTIPLE CORRELATION

15.1. The product-moment coefficient of correlation can, as has been seen in the last chapter, be used to measure the relationship between two variates which are distributed either exactly or approximately in the normal form. When we come to interpret such a correlation, however, we meet the same sort of problem which arose in Chapter 13 in connection with associations: if a variate 1 is correlated with a variate 2, may this not be due to the fact that both are correlated with a variate 3? The question may be decided by considering the correlation of 1 and 2 in the sub-populations for which variate 3 is constant, and in this chapter we consider the theory of such *partial* correlations, which bear an obvious analogy to the partial associations of Chapter 13. The subject may best be broached by extending to several variables the theory of linear regression developed for two variables in the previous chapter.

15.2. Suppose, in fact, that there is given a set of N individuals considered according to p variates $x_1, x_2, \ldots x_p$, so that to each individual there correspond p variate-values. We may, for example, be given a set of men according to height, weight, age and income, or a set of counties according to wheat-yields, hours of sunshine per annum, inches of rainfall per annum, and mean height above sea-level. In general, any variate may be considered as dependent on the others and for any variate, say x_1, we may require to find the "best" linear relation of the form

$$X_1 = \alpha + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_p X_p \quad . \quad . \quad . \quad . \quad (15.1)$$

a generalisation of (14.8). As before, the constants may be determined by the principle of least squares, i.e. so that

$$U = \Sigma(x_1 - \alpha - \beta_2 x_2 - \ldots - \beta_p x_p)^2 \quad . \quad . \quad . \quad . \quad (15.2)$$

is a minimum, the summation extending over the N members of the population. We shall then have

$$\frac{\partial U}{\partial \alpha} = \Sigma(x_1 - \alpha - \beta_2 x_2 - \ldots - \beta_p x_p) = 0, \qquad (15.3)$$

and if we take the variables measured from their means, this reduces to $\alpha = 0$. With this convention we have $(p - 1)$ equations of type $\dfrac{\partial U}{\partial \beta_k} = 0$, i.e.

$$x_k(x_1 - \beta_2 x_2 - \ldots - \beta_p x_p) = 0$$

or

$$\text{cov}\,(x_k, x_1) - \beta_2 \,\text{cov}\,(x_k, x_2) \qquad - \beta_k \,\text{var}\, x_k \qquad \ldots - \beta_p \,\text{cov}\,(x_k x_p) = 0,$$
$$k = 2, 3, \ldots p. \quad . \quad . \quad (15.4)$$

These $(p - 1)$ equations can be solved for the $(p - 1)$ quantities β and hence the required form (15.1) is determinate.

15.3. In the notation introduced by Yule we write

$$X_1 = \beta_{12.34\ldots p} X_2 + \beta_{13.24\ldots p} X_3 + \ldots + \beta_{1p.23\ldots(p-1)} X_p, \quad . \quad (15.5)$$

which is the regression equation of X_1 on $X_2 \ldots X_p$, referred to the means of the variates. The quantities β are called Partial Regression Coefficients. The first subscript to the left of the period in each β is that of the variate on the left of the regression equation, and the second subscript is that of the variate to which it is attached. These are called Primary Subscripts. The subscripts on the right of the period are those of the remaining variables and are called Secondary Subscripts.

When no confusion is likely to arise we can write (15.5) in the simpler form

$$X_1 = \beta_2 X_2 + \ldots + \beta_p X_p, \quad . \quad . \quad . \quad . \quad . \quad (15.6)$$

that is to say, we may drop the first primary and the secondary subscripts.

The order of the primary subscripts is material, $\beta_{12.k}$ being different from $\beta_{21.k}$; but that of the secondary subscripts is not.

Write

$$x_{1.23\ldots p} = x_1 - \beta_{12.34\ldots p}x_2 - \ldots - \beta_{1p.23\ldots(p-1)}x_p. \quad . \quad . \quad (15.7)$$

$x_{1.23\ldots p}$ may then be called the residual of x of order p. It is the difference between the observed x_1 and the value given by the regression equation. If all the residuals are zero, and only in this case, the regression is exactly linear. The β's were determined so as to make the sum of squares of residuals a minimum.

Write also

$$\text{var}\,(x_{1.23\ldots p}) = \sigma^2_{1.23\ldots p} \quad (15.8)$$

so that $\sigma_{1.23\ldots p}$ is the standard deviation of residuals and corresponds to the standard deviation of arrays considered in **14.22**.

15.4. From (15.7), equations (15.4) may be written

$$\Sigma(x_k x_{1.23\ldots p}) = 0, \; k = 2 \ldots p \quad . \quad . \quad . \quad . \quad . \quad (15.9)$$

and generally we shall have

$$\Sigma(x_k x_{j.12\ldots(j-1)(j+1)\ldots p}) = 0, \qquad j \neq k, \quad . \quad . \quad . \quad (15.10)$$

i.e. the covariance of any residual and any variate is zero, provided that the subscript of the latter occurs among the secondary subscripts of the former.

More generally still,

$$\Sigma(x_{1.34\ldots p}x_{2.34\ldots p}) = \Sigma\{x_{1.34\ldots p}(x_2 - \beta_{23.4\ldots p}x_3 - \ldots - \beta_{2p.34\ldots(p-1)}x_p$$

and each term on the right vanishes in virtue of (15.9) except the first, so that

$$\Sigma(x_{1.34\ldots p}x_{2.34\ldots p}) = \Sigma(x_{1.34\ldots p}x_2) \quad . \quad . \quad (15.11)$$

$$= \Sigma(x_1 x_{2.34\ldots p}) \quad . \quad . \quad (15.12)$$

by symmetry.

Thus the covariance of any two residuals is unaltered by omitting any or all of the secondary subscripts of either which are common to both. Conversely the covariance of any residual with p secondary suffixes and a residual with those p secondary suffixes and q additional ones is unaltered by adding to the former any of the q of the latter.

As a corollary, any covariance is zero if all the subscripts of one residual occur among the secondary subscripts of the second.

15.5. In virtue of these results we have

$$\begin{aligned} 0 &= \Sigma(x_{2.34\ldots p}x_{1.23\ldots p}) \\ &= \Sigma\{x_{2.34\ldots p}(x_1 - \beta_{12.34\ldots p}x_2 - \ldots)\} \\ &= \Sigma(x_{2.34\ldots p}x_1) - \beta_{12.34\ldots p}\Sigma(x_{2.34\ldots p}x_2) \\ &= \Sigma(x_{2.34\ldots p}x_{1.34\ldots p}) - \beta_{12.34\ldots p}\Sigma(x_{2.34\ldots p})^2. \end{aligned}$$

and thus, writing q for the group of suffixes $34 \ldots p$, we have

$$\beta_{12.q} = \frac{\operatorname{cov}(x_{1.q}, x_{2.q})}{\operatorname{var}(x_{2.q})} \tag{15.13}$$

a generalisation of (14.6).

Similarly

$$\beta_{21.q} = \frac{\operatorname{cov}(x_{1.q}, x_{2.q})}{\operatorname{var}(x_{1.q})} \tag{15.14}$$

We may then define a coefficient $\rho_{12.q} = \rho_{12.34\ldots p}$ by the equation

$$\begin{aligned}\rho_{12.q} &= (\beta_{12.q}\,\beta_{21.q})^{\frac{1}{2}} \\ &= \frac{\operatorname{cov}(x_{1.q}, x_{2.q})}{\{\operatorname{var}(x_{2.q})\operatorname{var}(x_{1.q})\}^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad (15.15)\end{aligned}$$

This is a generalisation of (14.10). $\rho_{12.q}$ is evidently the product–moment coefficient of correlation between $x_{1.q}$ and $x_{2.q}$.

15.6. From p variates we can pick out two in $\binom{p}{2}$ ways and find the regression of each on the other and their correlation; we can also pick out three in $\binom{p}{3}$ ways and find the regression of each on the other two; and so on. The number of possible regressions and correlations is thus very large, but they can all be expressed in terms of the variances of the variates and the correlations between pairs.

We shall call the coefficients with k secondary subscripts regressions, correlations, etc., of the kth order. The correlation between a pair of variates ρ_{ik} is thus of zero order, and our result may be stated in the form that coefficients of any order are expressible in terms of those of zero order. The proof follows from the expressions which we proceed to derive, giving coefficients of any order in terms of those of lower order. We have

$$\begin{aligned}\Sigma(x^2_{1.23\ldots p}) &= \Sigma(x_{1.23\ldots(p-1)}\,x_{1.23\ldots p}) \\ &= \Sigma\{x_{1.23\ldots(p-1)}\,(x_1 - \beta_{1p.23\ldots(p-1)}x_p - \text{terms in } x_2 \text{ to } x_{(p-1)})\} \\ &= \Sigma(x^2_{1.23\ldots(p-1)}) - \beta_{1p.23\ldots(p-1)}\Sigma(x_{1.23\ldots(p-1)}\,x_{p.23\ldots(p-1)})\,;\end{aligned}$$

hence, dividing by N,

$$\begin{aligned}\operatorname{var}(x_{1.23\ldots p}) &= \operatorname{var}(x_{1.23\ldots(p-1)}) - \beta_{1p.23\ldots(p-1)}\,\beta_{p1.23\ldots(p-1)}\operatorname{var}(x_{1.23\ldots(p-1)}) \\ &= \operatorname{var}(x_{1.23\ldots(p-1)})\,(1-\rho^2_{1p.23\ldots(p-1)}) \quad . \quad . \quad . \quad . \quad . \quad (15.16)\end{aligned}$$

which may be regarded as a generalisation of (14.11). By continuing the process we have

$$\operatorname{var}(x_{1.23\ldots p}) = \operatorname{var}(x_1)(1-\rho^2_{12})(1-\rho^2_{13.2})(1-\rho^2_{14.32}) \ldots (1-\rho^2_{1p.23\ldots(p-1)}) \quad . \tag{15.17}$$

or

$$\frac{\sigma^2_{1.23\ldots p}}{\sigma^2_1} = (1-\rho^2_{12})(1-\rho^2_{13.2}) \ldots (1-\rho^2_{1p.23\ldots(p-1)}) \tag{15.18}$$

The subscripts of the ρ's can be eliminated in a different order, giving alternative forms such as

$$\frac{\sigma^2_{1.23\ldots p}}{\sigma^2_1} = (1-\rho^2_{13})(1-\rho^2_{14.3})(1-\rho^2_{12.34}) \ldots \text{etc.}$$

Thus the variance of a residual of order $p-1$ is expressible in variances of zero order and correlations of order $p-2$.

15.7. Equations (15.4) may be written

$$\rho_{12}\,\sigma_1\sigma_2 - \beta_{12.34\ldots p}\,\sigma_2^2 - \beta_{13.2\ldots p}\,\rho_{23}\,\sigma_2\sigma_3 - \ldots = 0$$
$$\rho_{13}\,\sigma_1\sigma_3 - \beta_{12.34\ldots p}\,\rho_{23}\,\sigma_2\sigma_3 - \beta_{13.2\ldots p}\,\sigma_3^2 - \ldots = 0$$

etc. Adding the expression for $\Sigma(x^2_{1.23\ldots p})$, i.e.

$$\sigma_1^2 - \sigma^2_{1.23\ldots p} - \beta_{12.34\ldots p}\,\rho_{12}\,\sigma_1\sigma_2 - \beta_{13.2\ldots p}\,\rho_{13}\,\sigma_1\sigma_3 - \ldots = 0$$

we have p equations from which, on elimination of the β's, there results

$$\begin{vmatrix} \sigma_1^2 - \sigma^2_{1.23\ldots p} & \rho_{12}\,\sigma_1\sigma_2 & \rho_{13}\,\sigma_1\sigma_3 & \cdots \\ \rho_{21}\,\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\,\sigma_2\sigma_3 & \cdots \\ \rho_{31}\,\sigma_3\sigma_1 & \rho_{32}\,\sigma_3\sigma_2 & \sigma_3^2 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix}$$

where, of course, $\rho_{ik} = \rho_{ki}$. Dividing the ith row by σ_i and the kth column by σ_k, we get

$$\begin{vmatrix} 1 - \dfrac{\sigma^2_{1.23\ldots p}}{\sigma_1^2} & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \rho_{23} & & \rho_{2p} \\ \cdot & \cdot & \cdot & & \cdot \\ \rho_{1p} & \rho_{2p} & \rho_{3p} & & 1 \end{vmatrix} = 0 \qquad . \quad (15.19)$$

Write

$$\omega = \begin{vmatrix} 1 & \rho_{12} & \rho_{13} & & \rho_{1p} \\ \rho_{12} & 1 & \rho_{23} & & \rho_{2p} \\ & \cdot & \cdot & & \cdot \\ \rho_{1p} & \rho_{2p} & \rho_{3p} & & 1 \end{vmatrix} \qquad . \quad (15.20)$$

and ω_{11} for the minor of the first row and column of this determinant. Then from (15.19)—

$$\omega - \frac{\sigma_{1.23\ldots p}}{\sigma_1^2}\omega_{11} = 0$$

$$\sigma^2_{1.23\ldots p} = \frac{\omega}{\omega_{11}} \qquad . \quad (15.21)$$

Generally it may be shown in exactly the same way that

$$\text{cov}\,(x_{l.1\ldots l-1,\,l+1,\ldots p}\;x_{m.1\ldots m-1,\,m+1,\ldots p})$$
$$\frac{\sigma_l\sigma_m\omega}{\omega_{lm}} \qquad . \quad (15.22)$$

where ω_{lm} is the minor of the lth row and mth column in (15.20).

This result shows that the variances and covariances of residuals of any order can be expressed in terms of the correlations and variances of zero order.

15.8. We have, as in (15.16),

$$\Sigma(x_{1.34\ldots p}\,x_{2.34\ldots p}) = \Sigma(x_{1.34\ldots(p-1)}\,x_{2.34\ldots(p-1)}) - \beta_{2p.34\ldots(p-1)}\,\Sigma(x_{1.34\ldots(p-1)}\,x_{p.34\ldots(p-1)}).$$

Substituting

$$\beta_{2p.34\ldots(p-1)} = \beta_{p2.34\ldots(p-1)}\frac{\sigma^2_{2.34\ldots(p-1)}}{\sigma^2_{p.34\ldots(p-1)}}$$

and expressions for the covariances in terms of variances and regressions, and writing q for the group of secondary suffixes $34 \ldots (p-1)$, we find

$$\beta_{12.qp}\,\sigma^2_{2.qp} = \beta_{12.q}\,\sigma^2_{2.q} - \beta_{1p.q}\,\beta_{p2.q}\,\sigma^2_{2.q}.$$

whence, in virtue of (15.15),

$$\beta_{12.qp} = \frac{\beta_{12.q} - \beta_{1p.q}\,\beta_{p2.}}{1 - \beta_{2p.q}\,\beta_{p2.q}} \tag{15.23}$$

expressing the partial regression coefficient in terms of those of next lower order.

Writing down the similar equation for $\beta_{21.qp}$ and taking square roots, we find

$$\rho_{12.qp} = \frac{\rho_{12.q} - \rho_{1p.q}\,\rho_{2p.q}}{\{(1 - \rho^2_{1p.q})(1 - \rho^2_{2p.q})\}^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \tag{15.24}$$

a fundamental equation giving the correlation coefficient in terms of those of lower orders.

15.9. From the above results it is clear that the whole complex of partial regressions, correlations and variances or covariances of residuals is completely determined by the variances and correlations, or by the variances and regressions, of zero order. It is interesting to consider this result from the geometrical point of view.

Suppose in fact that we have N sets of observations of p variates

$$x_{11} \qquad x_{1p},\ x_{21} \qquad x_{2p}, \qquad x_{N1} \qquad x_{Np}.$$

Consider a Euclidean (flat) space of N dimensions. To each set of values $x_{1k}\ .\ .\ .\ x_{Nk}$ there will correspond one point in this space, and the totality of points representing all observations will be p in number. (This method of representation, it should be noted, is not that of N points in a p-way space, which was the one used in some of the sampling discussions in Chapter 10.) Call these points $Q_1, Q_2, \ .\ .\ .\ Q_p$. We will assume that the x's are measured about their mean, and take the origin to be P.

The quantity $N\sigma_l^2$ may then be interpreted as the square of the length of the vector joining the point $Q_l(= x_{1l},\ .\ .\ .\ x_{Nl})$ to P. Similarly ρ_{lm} may be interpreted as the cosine of the angle $Q_l P Q_m$, for

$$\rho_{lm} = \frac{\Sigma(x_{jl}\,x_{jm})}{(\Sigma x^2_{jl}\,\Sigma x^2_{jm})^{\frac{1}{2}}},$$

which is the formula for the cosine of the angle between PQ_l and PQ_m.

Our result may then be expressed by saying that all the relations connecting the p points in the N-space are expressible in terms of the lengths of the vectors OQ and the angles between them ; and the theory of partial correlation and regression is thus exhibited as formally identical with the trigonometry of an N-dimensional constellation of points.

15.10. The reader who prefers the geometrical way of looking at this branch of the subject will have no difficulty in translating the foregoing equations into trigonometrical terminology. We will here indicate only the more important results required for later sampling investigations.

Note in the first place that the p points Q and the point P determine (except perhaps in degenerate cases) a space of p dimensions in the N-space. Consider the point $Q_{1.2\ldots p}$ whose co-ordinates are the N residuals $x_{1.2\ldots p}$. In virtue of (15.9) the vector $PQ_{1.2\ldots p}$ is orthogonal to each of the vectors $PQ_2\ .\ .\ .\ PQ_p$ and hence to the space of $(p-1)$ dimensions defined by $P, Q_2, \ .\ .\ .\ Q_p$.

Consider now the residual vectors $Q_{1.q}$, $Q_{2.q}$, where q represents the secondary suffixes $34\ .\ .\ .\ (p-1)$. The cosine of the angle between them, say θ, is $\rho_{12.q}$ and each is orthogonal to the space $P, Q_3\ .\ .\ .\ Q_{(p-1)}$. Now take M on PQ_p such that $MQ_{1.q}$ and $MQ_{2.q}$ are perpendicular to PQ_p. Then $MQ_{1.q}$ is perpendicular to the space $P, Q_3, \ .\ .\ .\ Q_p$ and

so is $MQ_{2.q}$. The cosine of the angle between them, say ϕ, is $\rho_{12.qp}$ (cf. Fig. 15.1). Thus, to express $\rho_{12.qp}$ in terms of $\rho_{12.q}$ we have to express ϕ in terms of θ, or the angle between

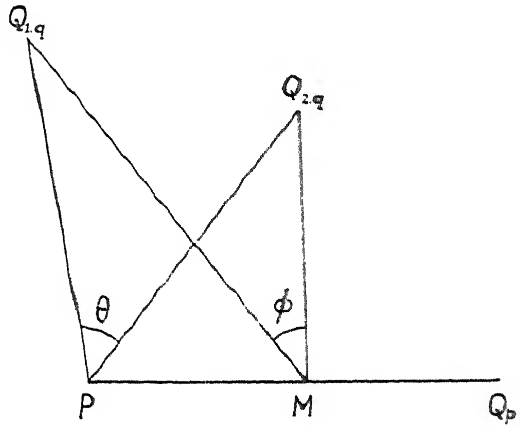


FIG. 15.1.

the vectors $PQ_{1.q}$ and $PQ_{2.q}$ in terms of that between their projections on the hyperplane perpendicular to PQ_p. We have

$$(Q_{1.q}\,Q_{2.q})^2 = PQ_{1.q}^2 + PQ_{2.q}^2 - 2PQ_{1.q}\,PQ_{2.q}\cos\theta$$
$$= MQ_{1.q}^2 + MQ_{2.q}^2 - 2MQ_{1.q}\,MQ_{2.q}\cos\phi.$$

Further
$$PQ_{1.q}^2 = PM^2 + MQ_{1.q}^2$$
and
$$PQ_{2.q}^2 = PM^2 + MQ_{2.q}^2$$
and hence we find
$$MQ_{1.q}\,MQ_{2.q}\cos\phi = -PM^2 + PQ_{1.q}\,PQ_{2.q}\cos\theta$$
$$\frac{MQ_{1.q}}{PQ_{1.q}}\frac{MQ_{2.q}}{PQ_{2.q}}\cos\phi = \cos\theta - \frac{PM}{PQ_{1.q}}\frac{PM}{PQ_{2.q}} \quad . \quad (15.25)$$

Now $\dfrac{MQ_{1.q}}{PQ_{1.q}}$ is the sine of the angle $MPQ_{1.q}$, the cosine of which angle is $\rho_{1p.q}$. Substituting in (15.25) we find

$$\cos\phi = \frac{\cos\theta - \rho_{1p.q}\,\rho_{2p.q}}{\{(1-\rho_{1p.q}^2)(1-\rho_{2p.q}^2)\}^{\frac{1}{2}}} \quad . \quad . \quad . \quad (15.26)$$

which is equation (15.24) in a slightly different form. The expression of a correlation coefficient in terms of those of the next lowest order is thus capable of interpretation as the projection of an angle on to a space of one fewer dimensions.

Example 15.1

In an investigation into the relationship between weather and crops, Hooker (1907) found the following means, standard deviations and correlations between the yields of seeds hay (x_1) in cwts. per acre, the spring rainfall (x_2) in inches and the accumulated temperature above 42° F. in the spring (x_3) for an English area over 20 years:—

$$\bar{x}_1 = 28{\cdot}02 \qquad \sigma_1 = 4{\cdot}42 \qquad \rho_{12} = -0{\cdot}80$$
$$\bar{x}_2 = 4{\cdot}91 \qquad \sigma_2 = 1{\cdot}10 \qquad \rho_{13} = -0{\cdot}40$$
$$\bar{x}_3 = 594 \qquad \sigma_3 = 85 \qquad \rho_{23} = -0{\cdot}56$$

The question of primary interest here is the influence of weather on crop yields, and we consider only the regression of x_1 on the other two variates. From the correlations of zero order it appears that yield and rainfall are positively correlated but that yield and accumulated spring temperature are negatively correlated. The question is, what interpretation is to be placed on this latter result? Does high temperature adversely affect yields or may the negative correlation be due to the fact that high temperature involves less rain, so that the beneficial effect of warmth is more than offset by the harmful effect of drought?

To decide this question, let us calculate the partial correlations and regressions. From (15.24) we have

$$\begin{aligned}\rho_{12.3} &= \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}\\ &= \frac{0{\cdot}80 - (-0{\cdot}40)(-0{\cdot}56)}{\sqrt{\{1-(0{\cdot}40)^2\}\{1-(0{\cdot}56)^2\}}}\\ &= 0{\cdot}759.\end{aligned}$$

Similarly

$$\begin{aligned}\rho_{13.2} &= 0{\cdot}097\\ \rho_{23.1} &= -0{\cdot}436.\end{aligned}$$

We next require the regressions β and the variances of residuals. From (15.14) we have

$$\begin{aligned}\beta_{12.3} &= \frac{\operatorname{cov}(x_{1.3}\,x_{2.3})}{\operatorname{var} x_{2.3}}\\ &= \frac{\rho_{12.3}\,\sigma_{1.3}}{\sigma_{2.3}}.\end{aligned}$$

This, however, involves the calculation of $\sigma_{1.3}$ and $\sigma_{2.3}$ which are not in themselves of interest. We can obviate the process by noting that from (15.16)

$$\begin{aligned}\sigma_{1.23} &= \sigma_{1.3}(1-\rho_{12.3}^2)^{\frac{1}{2}}\\ \sigma_{2.13} &= \sigma_{2.3}(1-\rho_{12.3}^2)^{\frac{1}{2}}\end{aligned}$$

so that

$$\beta_{12.3} = \frac{\rho_{12.3}\,\sigma_{1.23}}{\sigma_{2.13}}.$$

The standard deviations $\sigma_{1.23}$ and $\sigma_{2.13}$ are of some interest and may be calculated from (15.18). We have

$$\begin{aligned}\sigma_{1.23} &= \sigma_1(1-\rho_{12}^2)^{\frac{1}{2}}(1-\rho_{13.2}^2)^{\frac{1}{2}}\\ &= \sigma_1(1-\rho_{13}^2)^{\frac{1}{2}}(1-\rho_{12.3}^2)^{\frac{1}{2}}\end{aligned}$$

the two forms offering a check on each other.

From the first we have

$$\begin{aligned}\sigma_{1.23} &= 4{\cdot}42\{1-(0{\cdot}8)^2\}^{\frac{1}{2}}\{1-(0{\cdot}097)^2\}^{\frac{1}{2}}\\ &= 2{\cdot}64.\end{aligned}$$

Similarly

$$\begin{aligned}\sigma_{2.13} &= 0{\cdot}594\\ \sigma_{3.12} &= 70{\cdot}1.\end{aligned}$$

Thus

$$\beta_{12.3} = \frac{(0{\cdot}759)(2{\cdot}64)}{0{\cdot}594} = 3{\cdot}37,$$

and we also find

$$\beta_{13.2} = 0{\cdot}00364.$$

The regression equation of X_1 on X_2 and X_3 is then

$$X_1 - 28{\cdot}02 = 3{\cdot}37(X_2 - 4{\cdot}91) + 0{\cdot}0364(X_3 - 594).$$

This equation shows that for increasing rainfall the yield increases and that for increasing temperature the yield also increases, *other things being equal.* It enables us to isolate the effects of rainfall from those of temperature and study each separately. The positive regression $\beta_{13.2}$ means that there is a positive relation between yield and temperature when the effect of rainfall is eliminated. The partial correlations tell the same story. Although ρ_{13} is negative, $\rho_{13.2}$ is positive (though small), indicating that the negative value of ρ_{13} is due to complications introduced by the rainfall factor.

The foregoing procedure avoids the use of determinantal arithmetic, but the reader who prefers to do so may use equations (15.21). For example:

$$\omega = \begin{vmatrix} 1 & 0{\cdot}80 & -0{\cdot}40 \\ 0{\cdot}80 & 1 & -0{\cdot}56 \\ -0{\cdot}40 & -0{\cdot}56 & 1 \end{vmatrix}$$

$$= 0{\cdot}2448$$

$$\omega_{11} = \begin{vmatrix} 1 & -0{\cdot}56 \\ -0{\cdot}56 & 1 \end{vmatrix}$$

$$= 0{\cdot}6864,$$

from which

$$\sigma_{1.23} = \sigma_1\sqrt{\frac{\omega}{\omega_{11}}} = 2{\cdot}64 \text{ as before.}$$

15.11. When the work involves more than three variables it is desirable to systematise the arithmetic. Considerable assistance may be derived from tables of quantities such as

$$1 - \rho^2, \ \sqrt{(1 - \rho^2)}, \ \frac{1}{\sqrt{(1 - \rho_{12}^2)(1 - \rho_{13}^2)}}.$$

Kelley (1916, 1938) and Miner (1922) have given tables for this purpose. Trigonometrical tables are also useful in some cases. For instance, given ρ we can find $\theta = \cos^{-1}\rho$ and hence $\sin\theta\ (= \sqrt{(1 - \rho^2)})$, $\operatorname{cosec}\theta \left(= \dfrac{1}{\sqrt{(1 - \rho^2)}}\right)$, etc.

For determinantal work some systematic method of reduction such as the Doolittle method is useful.

Example 15.2

In some investigations into the variation of crime among cities in the U.S.A., Ogburn (1935) found a correlation of $-0{\cdot}14$ between crime rate (X_1) as measured by the number of known offences per thousand inhabitants and church membership (X_5) as measured by the number of church members of 13 years of age or over per 100 of total population of 13 years of age or over. The obvious inference is that religious belief acts as a deterrent to crime. Let us consider this more closely.

If X_2 = percentage of male inhabitants,
X_3 = percentage of total inhabitants who are foreign-born males, and
X_4 = number of children under 5 years old per 1000 married women between 15 and 44 years old.

[redacted]

Ogburn finds-

$$\begin{aligned} \rho_{12} &= +0{\cdot}44 & \rho_{24} &= -0{\cdot}19 \\ \rho_{13} &= -0{\cdot}34 & \rho_{25} &= -0{\cdot}35 \\ \rho_{14} &= -0{\cdot}31 & \rho_{34} &= +0{\cdot}44 \\ \rho_{15} &= -0{\cdot}14 & \rho_{35} &= +0{\cdot}33 \\ \rho_{23} &= +0{\cdot}25 & \rho_{45} &= -0{\cdot}85. \end{aligned}$$

From this and other data given in his paper it may be shown that we have, for the regression of X_1 on the other four variates,

$$X_1 - 19{\cdot}9 = 4{\cdot}51(X_2 - 49{\cdot}2) - 0{\cdot}88(X_3 \quad 30{\cdot}2) - 0{\cdot}072(X_4 - 4814) + 0{\cdot}63(X_5 - 41{\cdot}6),$$

and for certain partial correlations

$$\begin{aligned} \rho_{15.2} &= -0{\cdot}03 \\ \rho_{15.4} &= +0{\cdot}25 \\ \rho_{15.24} &= +0{\cdot}23. \end{aligned}$$

Now we note from the regression equation that when the other factors are constant X_1 and X_5 are positively related, i.e. church membership appears to be positively associated with crime. How does this effect come to be masked so as to give a negative correlation in the coefficient of zero order ρ_{15}?

We note in the first place that the correlation between crime and church membership when the effect of x_3, the percentage of foreigners, is excluded, is near zero. The correlation when x_4, the number of young children, is excluded, is positive; and the correlation when both x_3 and x_4 are excluded is again positive. It appears in fact from the regression equation that a high percentage of foreigners and a high proportion of children act as deterrents to crime. Now both these factors are positively correlated with church membership (foreign immigrants being mainly Catholic and more fecund). These correlations submerge the positive influence on crime of church membership among other members of the population. The apparently negative effect of church membership appears to be due to the more law-abiding spirit of the foreign immigrants and the fact that they are also more zealous churchmen.

The reader may care to refer to Ogburn's paper for a more complete discussion.

The Multivariate Normal Distribution

15.12. We now turn to consider the generalisation of the univariate and bivariate normal distributions to the case of p variables.

Consider the multivariate distribution

$$dF = y_0 \exp\left\{-\tfrac{1}{2}\sum_{r,s=1}^{p}\left(\alpha_{rs}\frac{x_r}{\sigma_r}\frac{x_s}{\sigma_s}\right)\right\}\frac{dx_1}{\sigma_1}\cdots\frac{dx_p}{\sigma_p} \qquad (15.27)$$

This has p variates and evidently reduces, when $p = 1$ or 2, to the normal type. We shall take it to be the generalisation of the normal distribution, and proceed to consider how the constants α are related to the correlations of the variates. It is, of course, assumed that the α's are such as to ensure the convergence of the distribution function. For this it is necessary and sufficient that the quadratic form $\Sigma\,\alpha_{rs}\frac{x_r}{\sigma_r}\cdot\frac{x_s}{\sigma_s}$ shall be positive-definite, i.e. that there is a real linear transformation reducing it to the sum of squares of p (or, in degenerate cases, fewer) new variates.

Make the transformation

$$\frac{x_r}{\sigma_r} = \sum_{j=1}^{p} l_{rj}\,\xi_j \quad . \quad . \quad . \quad . \qquad (15.28)$$

and choose the l's so that the exponent of (15.27) becomes $-\frac{1}{2}\Sigma\xi^2$. Then we have

$$\Sigma\,\alpha_{rs}\frac{x_r}{\sigma_r}\,\frac{x_s}{\sigma_s} = \Sigma\,\alpha_{rs}\,l_{rj}\,l_{sk}\,\xi_j\,\xi_k = \Sigma\xi^2$$

and hence, writing (α) for the matrix of the quantities α, (l) for that of the l's and $(\bar{l})$ for the transpose of (l), we have

$$(\alpha)(\bar{l})(l) = 1. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15.29)$$

Further, the Jacobian of the transformation is $|\,l\,|$, the determinant of the l's, and hence the integral of dF is given by

$$y_0\,|\,l\,|\int_{-\infty}^{\infty}\,.\,.\,.\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}\Sigma\,\xi^2\}\,d\xi_1\,.\,.\,.\,d\xi_p = (2\pi)^{\frac{p}{2}}y_0\,|\,l\,|.$$

Hence, since from (15.29) $|\,\alpha\,|\;|\,l\,|^2 = 1$, we have

$$y_0 = \frac{}{(2\pi)^{\frac{p}{2}}\,|\,l\,|} \qquad \frac{}{(2\pi)^{\frac{p}{2}}} \quad . \qquad (15.30)$$

Let us now find the characteristic function of the distribution. We have to integrate over the range of x's the exponential of

$$-\tfrac{1}{2}\left[\sum\left(\alpha_{rs}\frac{x_r}{\sigma_r}\,\frac{x_s}{\sigma_s}\right) - 2\sum\left(it_r\sigma_r\,\frac{x_r}{\sigma_r}\right)\right]$$
$$= -\tfrac{1}{2}[\Sigma(\xi_s^2) - 2\Sigma(it_r\sigma_r\,l_{rj}\,\xi_j)]$$
$$= -\tfrac{1}{2}[\Sigma(\xi_s - \Sigma\,it_r\sigma_r\,l_{rs})^2 + \sum_{j,k}[(t_jt_k\;\sigma_j\sigma_k\,\Sigma l_{jr}\,l_{kr})].$$

The first part reduces on integration to a constant. The second gives the exponential of a series of terms of second degree in t, the coefficient of $t_jt_k\;\sigma_j\sigma_k$ being

$$-\tfrac{1}{2}\sum_r(l_{jr}\;l_{kr}).$$

Now $\Sigma\;l_{jr}\,l_{kr}$ is the minor of the jth row and kth column in the matrix $(\bar{l})(l)$ and hence, from (15.29), in the matrix $(\alpha)^{-1} = (A)$ say. Hence we may write

$$\phi(t_1,\,.\,.\,.\,t_p) = \exp\{-\tfrac{1}{2}\Sigma\,(A_{jk}\,\sigma_j\sigma_k\;t_jt_k)\}. \quad . \quad . \quad . \qquad (15.31)$$

But when this is expanded the term in $\sigma_j\sigma_k\,t_jt_k$ is $-\rho_{jk}$ by definition and hence (A) is the matrix (ω) of equation (15.20). Thus

$$\phi(t_1,\,.\,.\,.\,t_p) = \exp\{-\tfrac{1}{2}\Sigma(\rho_{jk}\;\sigma_j\sigma_k\,t_jt_k)\}. \quad . \quad . \quad . \qquad (15.32)$$

Furthermore,

$$(\alpha) = (A)^{-1} = (\omega)^{-1}$$

and hence the distribution itself may be written in the form

$$dF = \frac{1}{(2\pi)^{\frac{n}{2}}\,\omega^{\frac{1}{2}}}\,\exp\left\{-\frac{1}{2\omega}\,\Sigma\left(\omega_{rs}\frac{x_r}{\sigma_r}\,\frac{x_s}{\sigma_s}\right)\right\}\frac{dx_1}{\sigma_1} \qquad (15.33)$$

For example, with the bivariate form

and hence $\omega = 1 - \rho^2$, $\omega_{11} = \omega_{22} = 1$, $\omega_{12} = \omega_{21} = -\rho$, so that the distribution becomes the familiar form

$$dF = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}\frac{dx_1}{\sigma_1}\frac{dx_2}{\sigma_2}.$$

15.13. For any fixed $x_2 \ldots x_p$ the exponent of (15.33) reduces to the normal univariate form in x_1 with mean

$$\frac{\sigma_1}{\omega_{11}}\left(\omega_{12}\frac{x_2}{\sigma_2} + \omega_{13}\frac{x_3}{\sigma_3} + \ldots \omega_{1p}\frac{x_p}{\sigma_p}\right) \quad . \quad . \quad . \quad (15.34)$$

Thus the regression of x_1 on the other variates is exactly linear. The variance of x_1 in any array is $\frac{\omega\sigma_1^2}{\omega_1}$ and the distribution is thus homoscedastic. It follows generally that the regression of any variate on any or all of the others is linear. Comparing (15.33) with (15.21) we see that the distribution may be written

$$dF = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma_1 \ldots \sigma_p\, \omega^{\frac{1}{2}}} \exp\left\{\frac{1}{2}\Sigma\frac{\rho_{rs.12\ldots p}\, x_r x_s}{\sigma_{r.12\ldots p}\,\sigma_{s.12\ldots p}}\right\}dx_1 \ldots dx_n \quad (15.35)$$

where the secondary suffixes in the ρ and σ's do not, of course, contain r and s.

Since every x is normally distributed, every linear function of x is so, as may be seen at once from (15.33). In particular the residuals are normally distributed.

If in (15.33) we make the substitution

$$\begin{aligned} \zeta_1 &= x_1 \\ \zeta_2 &= x_{2.1} \\ \zeta_3 &= x_{3.21} \\ \zeta_4 &= x_{4.321} \text{ etc.} \end{aligned}$$

the exponent will be a quadratic function of the ζ's. In this function all product terms $\zeta_j\,\zeta_k, j \neq k$, must vanish, for the covariance of ζ_j and ζ_k vanishes in virtue of the remark at the end of **15.4.** It follows that the distribution function may be written in the form

$$dF = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma_1\,\sigma_{2.1}\,\sigma_{3.21}} \exp\left\{-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_{2.1}^2}{\sigma_{2.1}^2} + \frac{x_{3.21}^2}{\sigma_{3.21}^2} + \ldots\right)\right\}dx_1\,dx_{2.1}. \quad . \quad (15.36)$$

From this it appears that the joint distribution of any two residuals $x_{j.q}$ and $x_{k.q}$ is of the bivariate normal form with correlation $\rho_{jk.q}$. Consider, for example, $x_{2.1}$ and $x_{3.21}$. Each is normally distributed and is uncorrelated with and independent of the other variables in (15.36). If $x_{3.21}$ is expressed in terms of residuals of the second order, i.e. as $x_{3.1} - \beta_{32.1}x_{2.1}$, the joint distribution of $x_{2.1}$ and $x_{3.1}$ becomes of the bivariate form with correlation $\rho_{23.1}$; and so generally.

These results are important in the interpretation of regressions and correlations in the normal case. In the general case a coefficient such as $\rho_{jk.q}$ represents the average dependence, so to speak, of $x_{j.q}$ and $x_{k.q}$, being based on the sum $\Sigma(x_{j.q}\, x_{k.q})$. In the normal case $\rho_{jk.q}$ is constant for all the sub-populations corresponding to particular assigned values of the other variables.

Sampling Distributions of Partial Correlation and Regression Coefficients

15.14. We now consider the sampling distributions of the coefficients of partial correlation and regression. For large samples the values of Chapter 14 appropriate to

correlations and regressions of zero order may be used (subject to the proviso as to the unreliability of the standard error for ρ unless the sample is very large). For example, the variance of $\rho_{jk.q}$ in the normal case is given by

$$\operatorname{var}(r_{jk.q}) = \frac{1}{n}(1 - \rho^2_{jk.q})^2. \qquad (15.37)$$

where n is the sample number; and that of the regression coefficient by

$$\operatorname{var}(b_{jk.q}) = \frac{1}{n}\,\frac{\sigma^2_{j.kq}}{\sigma^2_{k.q}}. \qquad (15.38)$$

The proof of these results by the direct methods of Chapter 9 is a very tedious piece of algebra. They follow simply, however, from the remark of the previous section that the correlation between any two deviations $x_{j.q}$ and $x_{k.q}$ is of the normal type with coefficient $\rho_{jk.q}$; for it follows that $\rho_{jk.q}$ is distributed as the correlation between two normal variates. Similar considerations apply to the regression coefficients. It will be shown presently that if the original distribution was based on n observations, that of $\rho_{jk.q}$ is of the form of the correlation ρ_{jk} based on $n - s$ observations, where s is the number of secondary subscripts in q; but as our equations are only true to order n^{-1} the divisor in (15.37) and (15.38) may remain at n without further error.

15.15. Consider now the geometrical representation of **15.9.** Suppose we have three points Q, R, S in the n-fold space, represented by $x_1 \ldots x_n$, $y_1 \ldots y_n$, $z_1 \ldots z_n$ respectively, the origin being P and the variables measured from their mean. Then the coefficient of correlation between x and y is the cosine of the angle QPR, that between y and z the cosine of RPS and that between z and x the cosine of SPR. Now imagine a sphere described with unit radius and centre P, cutting PQ, PR and PS in Q', R', S'. Then will the partial correlation $r_{xy.z}$ be the cosine of the angle of the spherical triangle $Q'S'R'$, and so for the other two partial correlations. This was, in effect, proved in **15.10**, for the angle $Q'S'R'$ is the angle between the projections of PQ and PR upon the space perpendicular to PS.

Now we may make an orthogonal transformation, corresponding to a rotation of the co-ordinate axes, without affecting the correlations; moreover, if the n values of one variate x are independent and normally distributed so will be the n values of the transformed variates. Let us then make such a transformation and take PS as one of the new co-ordinate axes. It is then apparent that the distribution of $r_{xy.z}$, which is the cosine of an angle in the space perpendicular to PS, is the same in form as that of r_{xy} except that, being in $(n - 1)$ dimensions, it is based on $(n - 1)$ independent pairs of normally distributed variates instead of n.

Hence for samples from a normal population the distribution of the partial correlation coefficient of the first order from n sets of observations is the same as that of a correlation of zero order from $(n - 1)$ sets of observations. By a repetition of the same argument it follows that the distribution of a correlation coefficient of the sth order is that of the correlation of zero order from $(n - s)$ sets of observations. The results of the previous chapter are thus immediately applicable to partial correlations. If, of course, s is small compared with n, the distribution of partials is sensibly the same as that of ordinary correlations, which confirms the approximation of the previous section.

This can be reduced to the z-form by writing

$$\left.\begin{aligned} z &= \tfrac{1}{2}\log_e \frac{R^2}{1-R^2}\,\frac{N-p}{p-1} \\ \nu_1 = p-1, &\quad \nu_2 = N-p \end{aligned}\right\} \quad . \qquad (15.50)$$

The mean value of R^2 is the positive quantity $(p-1)/(N-1)$.

15.21. We proceed to find the distribution of R in samples from a normal multivariate population when $\boldsymbol{R}$ is not zero. Two preliminary remarks are necessary.

In the first place, any multivariate normal population can, by a linear transformation, be transformed to new variates which are normally distributed and independent. One such transformation has been given in **15.13.**

Secondly, any linear transformation leaves the multiple correlation coefficient invariant, that is to say, the coefficient between x_1 and $x_2 \ldots x_p$ is the same as that between x_1 and the transformed variables $\xi_2 \ldots \xi_p$. Referring to (15.43) we see that, apart from the constant σ_1^2, $R_{1(2\ldots p)}$ depends only on $\sigma^2_{1.2\ldots p}$, and since the regressions are chosen so as to minimise this quantity, the same minimum is reached whether we use the variables $x_2 \ldots x_p$ or the linearly related variables $\xi_2 \ldots \xi_p$. Conversely, if the correlation between x_1 and ξ_2 is a maximum for all possible sets of ξ's, then that correlation is the multiple correlation coefficient between x_1 and the ξ's, and x_1 is uncorrelated with $\xi_3 \ldots \xi_p$.

From the geometrical standpoint of **15.10,** let us take the sample vectors $PQ_2 \ldots PQ_p$ and in the space defined by these vectors choose another set $PS_2 \ldots PS_p$ which are mutually orthogonal. These will correspond to the transformed variates ξ, and the angle between PQ_1 and the space remains unaltered, i.e. R is invariant.

Let us now choose ξ_2 so that the correlation between x_1 and ξ_2 is a maximum in the population. Then if PS_2 is the sample vector corresponding to ξ_2, PQ_1 will be orthogonal to all the other vectors $PS_3 \ldots PS_p$ (since x_1 is then independent of $\xi_3 \ldots \xi_p$).

In any given sample value the correlation between x_1 and ξ_2 will not be equal, in general, to $\boldsymbol{R}$ (though the correlation in the population is $\boldsymbol{R}$), but to a quantity $\boldsymbol{r}$, say, varying from sample to sample and equal to $\cos^{-1}Q_1PS_2$. Let PT be the vector representing the sampling regression formula $\sum_{j=1}^{p} \nu_{ij}$. This will lie in the x-ξ space (cf. Fig. 15.2). Then

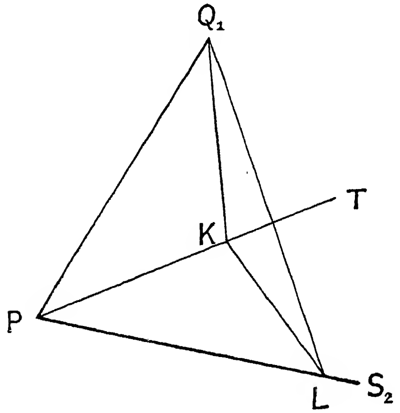


Fig. 15.2.

PT is such as to make the greatest possible angle with Q_1P, the angle being $\cos^{-1} R$, and the perpendicular from Q_1 on to the x-ξ space meets PT in a point, say K. Let the point L be taken on PS_2 such that LK is perpendicular to PK and join Q_1L. Let the angle KPL be ψ.

Then
$$\begin{aligned} Q_1L^2 &= Q_1P^2 + PL^2 - 2Q_1P.PL\,\boldsymbol{r} \\ &= Q_1K^2 + KL^2 \\ &= Q_1P^2 - PK^2 + PL^2 - PK^2 \end{aligned}$$

and hence

$$\boldsymbol{r} = \frac{PK^2}{Q_1P.PL} - \frac{PK}{Q_1P}\cdot\frac{PK}{PL}$$
$$= R\cos\psi. \qquad (15.51)$$

R and ψ are independent.

Now we consider the sampling distribution of the correlation coefficient $\boldsymbol{r}$. It is to be remembered that x_1 and ξ_2 are distributed in the bivariate normal form. The distribution of $\boldsymbol{r}$ is then given by the formulae of **14.14** and may be written

$$dF = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-\boldsymbol{r}^2)^{\frac{n-4}{2}}\,d\boldsymbol{r}$$

$$\times \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}(1-\boldsymbol{R}^2)^{\frac{n-1}{2}}\int_{-\infty}^{\infty}\frac{dz}{(\cosh z - \boldsymbol{R}\boldsymbol{r})^{n-1}} \qquad (15.52)$$

since $\boldsymbol{R}$ is the population value of $\boldsymbol{r}$. If $\boldsymbol{R} = 0$ the second factor in (15.52) reduces to unity. We may therefore regard the second factor as the effect on the frequency density in the region $d\boldsymbol{r}$ of a population correlation $\boldsymbol{R}$. Now we have already seen that when $\boldsymbol{R} = 0$ the distribution of R corresponding to that of $\boldsymbol{r}$ is given by (15.49). We have then to find by what factor (15.49) is to be multiplied to allow for the variable frequency density. Such factor is the second part on the right-hand side of (15.52) with $R\cos\psi$ substituted for $\boldsymbol{r}$ (as given by (15.51)) and integrated over the permissible domain of ψ.

It will be seen from Fig. 15.2 that for fixed P and S_2, T may vary in the space of $(p-1)$ dimensions determined by P and $S_2 \ldots S_p$; and for constant ψ it will lie on the cone in that space obtained by rotating TP about PS_2. The element of area cut off by this cone on the unit hypersphere is proportional to its solid angle, that is to $\sin^{p-3}\psi$, and hence the frequency of ψ in the range $d\psi$ is

$$\frac{\Gamma\left(\frac{p-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{p-2}{2}\right)}\sin^{p-3}\psi\,d\psi \qquad (15.53)$$

Thus the density factor is

$$\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}{\pi\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{p-2}{2}\right)}(1-\boldsymbol{R}^2)^{\frac{n-1}{2}}\int_0^{\pi}\int_{-\infty}^{\infty}\frac{\sin^{p-3}\psi\,dz\,d\psi}{(\cosh z - \boldsymbol{R}R\cos\psi)^{n-1}}. \qquad (15.54)$$

Finally the distribution of $\mathbf{R}$ is

$$dF = \frac{\Gamma\left(\frac{n}{2}\right)}{\pi\,\Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{p-2}{2}\right)}(1-\mathbf{R}^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-3}{2}}(1-R^2)^{\frac{n-p-2}{2}}d(R^2)$$

$$\times \int_0^{\pi}\int_{-\infty}^{\infty}\frac{\sin^{p-3}\psi\, dz\, d\psi}{(\cosh z - \mathbf{R}R\cos\psi)^{n-1}} \quad . \qquad (15.55)$$

This may be expressed as a hypergeometric function. Expanding the integrand in powers of $\cos\psi$ we have, since odd powers vanish on integration,

$$\sum_{j=0}\binom{n+2j-2}{2j}\frac{\sin^{p-3}\psi\cos^{2j}\psi}{(\cosh z)^{n-1+2j}}(\mathbf{R}R)^{2j}$$

and since $$\int_0^{\pi}\cos^{2j}\psi\sin^{p-3}\psi = B\left(\frac{p-2}{2},\frac{2j+1}{2}\right)$$

and $$\int_{-\infty}^{\infty}\frac{dz}{\cosh^{n-1+2j}z} = B\left(\frac{1}{2},\frac{n+2j-1}{2}\right)$$

the integral becomes

$$\Sigma\binom{n+2j-2}{2j}B\left(\frac{p-2}{2},\frac{2j+1}{2}\right)B\left(\frac{1}{2},\frac{n+2j-1}{2}\right)(\mathbf{R}R)^{2j}$$

$$= \frac{\pi\,\Gamma\left(\frac{p-2}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}F\left(\frac{n-1}{2},\frac{n-1}{2},\frac{p-1}{2},\mathbf{R}^2R^2\right) \qquad (15.56)$$

whence we find, from (15.55), after a little further reduction,

$$dF = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}(1-\mathbf{R}^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-3}{2}}(1-R^2)^{\frac{n-p-2}{2}}dR^2$$

$$\times F\left\{\frac{n-1}{2},\frac{n-1}{2},\frac{p-1}{2},\mathbf{R}^2R^2\right\}. \qquad (15.57)$$

15.22. Writing $a = \frac{1}{2}(p-1)$, $b = \frac{1}{2}(n-p)$ we have

$$dF = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(1-\mathbf{R}^2)^{a+b}(R^2)^{a-1}(1-R^2)^{b-1}F(a+b,a+b,a,\mathbf{R}^2R^2)\,dR^2 \qquad (15.58)$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{(1-\mathbf{R}^2)^{a+b}}{(1-\mathbf{R}^2R^2)^{a+2b}}(R^2)^{a-1}(1-R^2)^{b-1}\times F(-b,-b,a,\mathbf{R}^2R^2)\,dR^2 \quad . \qquad (15.59)$$

It may be shown that

$$\mu_1'(R^2) = 1 - \frac{b}{a+b}(1-\mathbf{R}^2)F(1,1,a+b+1,\mathbf{R}^2). \quad . \qquad (15.60)$$

In particular, when $\mathbf{R} = 0$ we have the known result

$$\mu_1'(R^2) = \frac{a}{a+b} = \frac{p-1}{n-p}. \quad . \quad . \quad . \quad . \quad . \qquad (15.61)$$

For large n we have approximately

$$\mu_1'(R^2) = \frac{a + (b - \frac{1}{2})\boldsymbol{R}^2 + \boldsymbol{R}^4}{a + b + \frac{1}{2}}. \qquad (15.62)$$

For the second moment

$$\mu_2(R^2) = \frac{b(b+1)(1-\boldsymbol{R}^2)^2}{(a+b)(a+b+1)} F(2, 2, a+b+2, \boldsymbol{R}^2)$$
$$- \frac{b^2(1-\boldsymbol{R}^2)^2}{(a+b)^2} F^2(1, 1, a+b+1, \boldsymbol{R}^2) \qquad (15.63)$$

or approximately

$$\mu_2(R^2) = \frac{4\boldsymbol{R}^2(1-\boldsymbol{R}^2)^2}{n} \qquad (15.64)$$

which, however, breaks down near $\boldsymbol{R} = 0$. It would, in fact, appear that the distribution of R tends to normality when $\boldsymbol{R} \neq 0$ but not when $\boldsymbol{R} = 0$ (cf. Exercise 15.3).

Example 15.3

From Example 15.1 we have found $\omega = 0.2448$, $\omega_{11} = 0.6864$, from which we have

$$R^2_{1(23)} = 1 - \frac{0{\cdot}2448}{0{\cdot}6864}$$
$$= 0{\cdot}6433,$$

indicating that the regression equation is a fairly close representation of the data, since R, the correlation between observed x_1's and those provided by the equation, is high, about 0.80.

It is hardly necessary to test the significance of such a value, but we will do so to illustrate the arithmetic involved. If x_1 were uncorrelated with the other variates we should have $\boldsymbol{R} = 0$, and on the assumption that the population is normal (a reasonable assumption for crop yields, sunshine and rainfall records) we may use equation (15.50). We have, since $p = 3$, $n = 20$

$$z = \tfrac{1}{2} \log_e \frac{0{\cdot}6433}{0{\cdot}3567} \cdot \frac{17}{2}$$
$$= 1{\cdot}36$$
$$\nu_1 = 2, \ \nu_2 = 17.$$

From Appendix Table 5 the 1 per cent. significance point of z for $\nu_1 = 2, \nu_2 = 17$ is $0{\cdot}9051$. so that the observed R is almost certainly significant, z being much greater than can be accounted for by sampling alone.

NOTES AND REFERENCES

The theory of partial correlation is mainly due to Yule (1907). The reader may refer to M. Ezekiel's book (1930) for a detailed discussion of the practical side of correlation analysis. See also a paper on the theoretical side by Frisch (1929).

For a knowledge of the sampling properties of the partial correlations we are indebted to Yule (1907), who pointed out the applicability of large sampling " normal " formulae for coefficients of zero order to the partial coefficients, and to R. A. Fisher (1924), who is responsible for the exact result for small samples from a normal population and the

distribution of the multiple correlation coefficient (1928). Some approximate results for the latter had been obtained by Isserlis (1917) and P. Hall (1927). Wishart (1931, 1932) has studied the exact distribution of R and the formally equivalent η. Both of Fisher's papers are notable examples of the power of the geometrical method of deducing sampling distributions.

In comparing formulae given by various writers it is as well to examine whether the total number of variates (our p) or the number of dependent variates $(p-1)$ is being used as a constant in the equations.

Ezekiel, M. (1930), *Methods of Correlation Analysis*, Chapman and Hall, London; John Wiley and Sons, New York.

Fisher, R. A. (1924), "The distribution of the partial correlation coefficient," *Metron*, **3**, 329.

—— (1928), "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc.*, A, **121**, 654.

Frisch, R. (1929), "Correlation and Scatter in statistical variables," *Nordic Statistical Journal*, **1**, 36.

Hall, P. (1927), "Multiple and partial correlation coefficients in the case of an n-fold variate system," *Biometrika*, **19**, 100.

Hooker, R. H. (1907), "The correlation of the weather and the crops," *Jour. Roy. Stat. Soc.*, **65**, 1.

Isserlis, L. (1914), "On the partial correlation ratio, Part I, Theoretical," *Biometrika*, **10**, 391, and "Part II, Numerical," *ibid.* (1916), **11**, 50.

—— (1917), "The variation of the multiple correlation coefficient in samples drawn from an infinite population with normal distribution," *Phil. Mag.*, **34**, 205.

Kelley, T. L. (1916), "Tables to facilitate the calculation of partial coefficients of correlation and regression equations," *Bulletin of the University of Texas*, No. 127.

—— (1938), *The Kelley Statistical Tables*, Macmillan.

Miner, J. R. (1922), "Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in partial correlations, etc.," Johns Hopkins Press, Baltimore.

Ogburn, W. F. (1935), "Factors in the variation of crime among cities," *Jour. Amer. Statist. Ass.*, **30**, 12.

Wishart, J. (1931), "The mean and second-moment coefficient of the multiple correlation coefficient in samples from a normal population," *Biometrika*, **22**, 353.

—— (1932), "Note on the correlation ratio," *Biometrika*, **23**, 441.

Yule, G. U. (1907), "On the theory of correlation for any number of variables treated by a new system of notation," *Proc. Roy. Soc.*, A, **79**, 182.

EXERCISES

15.1. Show that

$$\beta_{12.34\ldots(p-1)} = \frac{\beta_{12.34\ldots p} + \beta_{1p.23\ldots(p-1)}\,\beta_{p2.43\ldots(p-1}}{1 - \beta_{1p.23\ldots(p-1)}\,\beta_{p1.23\ldots(p-1)}}$$

and that

$$\rho_{12.34\ldots(p-1)} = \frac{\rho_{12.34\ldots p} + \rho_{1p.23\ldots(p-1)}\,\rho_{2p.13\ldots(p-1)}}{(1-\rho^2_{1p\ 23\ldots(p-1)})^{\frac{1}{2}}\,(1-\rho^2_{2p.13\ldots(p-1)})^{\frac{1}{2}}}$$

(Yule, 1907.)

15.2. Show that for p variates there are $\binom{p}{2}$ correlation coefficients of order zero and $\binom{p-2}{s}\binom{p}{2}$ of order s. Show further that there are $\binom{p}{2}2^{p-2}$ correlation coefficients altogether and $\binom{p}{2}2^{p-1}$ regression coefficients.

15.3. Show that for given ρ_{12}, ρ_{13}, ρ_{23} must lie in the range

$$\rho_{12}\,\rho_{13} \pm (1 - \rho_{12}^2 - \rho_{13}^2 + \rho_{12}^2\,\rho_{13}^2)^{\frac{1}{2}}$$

and that if x_1 and x_2, x_1 and x_3 are uncorrelated no inference can be drawn from that fact as to the correlation between x_2 and x_3.

15.4. Show that if ρ_{12} be zero, $\rho_{12.3}$ will not be zero unless at least one of ρ_{13}, ρ_{23} is zero.

15.5. If the correlations of zero order among a set of variables are all equal to ρ, show that every partial correlation of the sth order is $\dfrac{\rho}{(1 + s\rho)}$.

15.6. Show that the distribution of the multiple correlation coefficient R tends, in normal samples, for large n, to the form

$$dF = \frac{(\frac{1}{2}B^2)^{\frac{p-3}{2}}}{\Gamma\left(\frac{p-1}{2}\right)} \exp\{-\tfrac{1}{2}B^2 - \tfrac{1}{2}\beta^2\}$$
$$\times \left\{1 + \frac{1}{p-1}\frac{\beta^2 B^2}{2} + \frac{1}{(p-1)(p+1)}\frac{\beta^2 B^2}{2.4} + \dots\right\} d(\tfrac{1}{2}B^2)$$

where $\beta^2 = \mathbf{R}^2(n-p), \quad B^2 = R^2(n-p).$

In particular, where $p = 4$,

$$dF = \frac{1}{\sqrt{2\pi}}\frac{B}{\beta}[\exp\{-\tfrac{1}{2}(B-\beta)^2\} - \exp\{-\tfrac{1}{2}(B+\beta)^2\}]\,dB.$$

Thus, when $\beta = 0$ the distribution of B does not tend to normality, but when β is not zero and is thus large for finite $\mathbf{R}$, B is distributed approximately normally about β with variance unity.

(Fisher, 1928.)

15.7. Show that the distribution function of R in normal samples may be written, if $n - p$ is even, in the form

$$(1 - \mathbf{R}^2)^{\frac{n-1}{2}} R^{p-1} \sum_{j=0}^{\frac{1}{2}(n-p-2)} \frac{\Gamma\left(\frac{p-1+2j}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)j!} \frac{(1-R^2)^j}{(1-\mathbf{R}^2R^2)^{\frac{1}{2}(n-1+2j)}}$$
$$\times F\left\{-j, -\frac{n-p}{2}, \frac{p-1}{2}, \mathbf{R}^2R^2\right\}.$$

(Fisher, 1928.)

CHAPTER 16

RANK CORRELATION

16.1. In previous chapters we have considered the dependence of attributes, as measured by coefficients of association, and that of variables as measured (in the normal case at least) by product–moment correlation. In this chapter we shall consider a type of relationship which, in a sense, occupies an intermediate position between the two, the correlation of ranks.

Consider a set of individuals which can be arranged in order according to some quality, such as a set of men according to ability or a set of musical compositions according to the degree of preference with which they are regarded by some observer. An ordered arrangement of the objects will be called a ranking and the ordinal number of a given individual in the ranking is called his rank. Thus with a ranking of n individuals there will be one rank corresponding to each of the n ordinal numbers 1 to n.

16.2. Ranking is less general than the classification of attributes in the sense that the division of a population into classes A and not-A, or $A_1, A_2 \ldots A_p$, does not require any ordering of those classes; the measures of contingency and association discussed in Chapter 13 are invariant under rearrangements of columns or rows in the tables. On the other hand, individuals arranged in an ordinary frequency table have their interrelationships more closely defined than if they are merely ranked, so that ranking is in a sense more general than measurement according to a variate-scale. To put the point in a slightly different way, a ranking is invariant under any transformation which stretches the scale of measurement of the variate.

16.3. In practice, ranked data usually arise in two ways:—

(*a*) From material which could be measured on a variate-scale but which is not so measured for reasons of economy, lack of adequate instruments, and so forth. This class includes the case where the data are given as measurements but are then ranked on the basis of those measurements in order, for example, to reduce the arithmetical work in investigating correlations.

(*b*) From material which is believed to be capable of measurement theoretically but cannot be measured in practice, e.g. human preferences for food or intelligence. Ranking methods are sometimes applied rather uncritically to material which the experimenter considers to be capable of ranking, whether it has been demonstrated to be so or not. We shall return to this point below.

It is always possible by suitable conventions to impose a scale of measurement and hence a variate-system on ranked material; but the process is sometimes rather artificial and we shall in the first instance consider ranked material as such, without reference to the possibility of there being any pre-existent or superimposed variate in the background.

Spearman's Coefficient of Rank Correlation

16.4. Consider a set of n individuals ranked according to two variables in the orders $X_1, X_2, \ldots X_n$, $Y_1, Y_2 \ldots Y_n$, where the X's and the Y's are permutations of the

numbers 1 to n. Our problem is to discuss the relationship between the X's and the Y's. If the individuals are denoted by $A_1 \ldots A_n$ we may write the rankings in the form

Individual	A_1	A_2
Ranking 1	X_1	X_2
Ranking 2	Y_1	Y_2

(16.1)

We note first of all that the concordance between rankings is perfect if and only if $X_j = Y_j$ for all j. It is natural to consider the differences $X_j - Y_j$ ($= d_j$, say) as measuring the difference between the two rankings. They are zero if and only if the concordance is perfect and their magnitude to some extent reflects the divergence of the rankings from perfect concordance. We also note that

$$\sum_{j=1}^{n} d_j = \sum_{j=1}^{n} X_j - \sum_{j=1}^{n} Y_j = 0, \qquad (16.2)$$

for each of the sums of X and Y is the sum of the first n natural numbers. We might then take $\Sigma \mid d \mid$ as a measure of discordance, and a coefficient based on this quantity was in fact proposed by Spearman (1906). It is however subject to several disadvantages, similar to those attaching to the mean deviation, and a more suitable measure is obtained by using $\Sigma(d^2)$. It is easy to see that the maximum value possible for $\Sigma(d^2)$ is $\frac{(n^3 - n)}{3}$. For $\Sigma(d^2)$ is the greatest if the d's are as different as possible, i.e. if one ranking is the reverse of the other, so that the d's are $(n-1), (n-3) \ldots -(n-3), -(n-1)$, though not necessarily in that order. In this case

$$\begin{aligned}
\Sigma(X_jY_j) &= 1(n) + 2(n-1) + 3(n-2) + \ldots n\{n-(n-1)\} \\
&= 1\{(n+1)-1\} + 2\{(n+1)-2\} + \ldots n\{(n+1)-(n)\} \\
&= (n+1)\sum_{j=1}^{n} j - \sum_{j=1}^{n} j^2 \\
&= \frac{n(n+1)(n+2)}{6}
\end{aligned} \qquad (16.3)$$

Thus
$$\begin{aligned}
\Sigma(d^2) &= \Sigma(X^2) + \Sigma(Y^2) - 2\Sigma(XY) \\
&= \frac{n(n+1)(2n+1)}{3} - \frac{n(n+1)(n+2)}{3}
\end{aligned}$$

(16.4)

We then define

$$= 1 - \frac{6\Sigma(d^2)}{n^3 - n} \qquad (16.5)$$

as the Spearman coefficient of rank correlation. If the concordance between rankings is perfect $\Sigma(d^2) = 0$ and $\rho = 1$. If the discordance is perfect $\rho = -1$. In other cases ρ lies between these limits.

It is worth noticing that ρ is the product–moment coefficient of correlation between X and Y when we regard the ranks as variate-values. For we then have

$$\mu_1'(X) = \mu_1'(Y) = \frac{n+1}{2} \quad . \quad . \quad . \tag{16.6}$$

$$n \operatorname{var} X = n \operatorname{var} Y = \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2$$

$$= \frac{n^3 - n}{12} \quad . \quad . \tag{16.7}$$

$$n \operatorname{cov}(X, Y) = \Sigma(XY) - n\{\mu_1'(X)\}^2$$

$$= -\tfrac{1}{2}\Sigma(X - Y)^2 + \Sigma(X^2) - \frac{n(n+1)^2}{4}$$

$$= \frac{n^3 - n}{12} - \tfrac{1}{2}\Sigma(d^2),$$

so that the product–moment correlation coefficient of X and Y is

$$\frac{\left(\frac{n^3 - n}{12} - \tfrac{1}{2}\Sigma d^2\right)}{\left(\frac{n^3 - n}{12}\right)}$$

$$= 1 - \frac{6\Sigma(d^2)}{n^3 - n} = \rho.$$

16.5. There is an element of artificiality in the Spearman coefficient as defined which we must remove. The ranks are ordinal numbers and cannot without justification be operated on by the laws of cardinal arithmetic. For instance, if A_1 is ranked 4th and 8th by two observers, d_1 is $(4 - 8)$; but what does 4th minus 8th mean, and what significance is to be attached to its square? It is not entirely trivial to note that the necessary transition from ordinals to cardinals may be made without invoking a variate-scale. When we rank a member as r we mean that in the set of n, $(r - 1)$ members are ranked higher. This number $(r - 1)$ is a cardinal and in our particular example 4th minus 8th may be regarded as meaning that the difference of the number of members ranked higher by the two observers was 4.

Example 16.1

Two judges in a beauty contest rank the 10 competitors in the following order:

6	4	3	1	2	7	9	8	10	5
4	1	6	7	5	8	10	9	3	2

What is the rank correlation?

The differences between the ranks are

$$2 \quad 3 \quad -3 \quad -6 \quad -3 \quad -1 \quad -1 \quad -1 \quad 7 \quad 3$$

which sum to zero as they should.

Thus
$$\Sigma(d^2) = 4 + 9 + 9 + 36 + \text{etc.}$$
$$= 128$$
$$\rho = 1 - \frac{6 \cdot 128}{990} = 0.224.$$

This indicates some sort of concordance between the standards of the two judges, but not a very strong concordance.

Example 16.2

In the previous example there was no information about the "real" order of the competitors, and ρ merely served to measure the degree of agreement between judges. Consider, however, the following case, where an objective order is known: In a test for ability to distinguish shades of colour, ten discs were prepared ranging from light to dark red, and a subject was asked to arrange them in order. The true order, as determined by a colorimetric method, was

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10.$$

The order produced by the subject was

$$4,\ 7,\ 2,\ 10,\ 3,\ \qquad\qquad 5,\ 9.$$

What sort of a judge is he?

The differences are

$$-3,\ -5,\ 1,\ -6,\ 2,\ 0,\ -1,\ 7,\ 4,\ 1$$

and $$\Sigma(d^2) = 142,\ \rho = 0{\cdot}139.$$

The coefficient is low and we conclude that the observer was a poor judge.

An Alternative Coefficient

16.6. A second coefficient of rank correlation which has certain advantages may be obtained as follows: Consider again the ranking of the previous example

$$4\quad 7\quad 2\quad 10\quad 3\quad 6\quad 8\quad 1\quad 5\quad 9 \qquad . \quad . \quad . \quad . \quad (16.8)$$

Consider the order of the nine pairs of numbers obtained by taking the first number 4 with each succeeding number. The first pair, 4, 7, is in the correct order (in the sequence 1, 2, . . . 10) and we therefore allot it the score + 1. The second pair, 4, 2, is in the wrong order and we therefore score − 1. The nine scores will be found to be

$$+1 - 1 + 1 - 1 + 1 + 1 - 1 + 1 + 1, \text{ totalling } + 3.$$

Consider next the scores of the second number 7, with its eight succeeding numbers. They are

$$-1 + 1 - 1 - 1 + 1 - 1 - 1 + 1, \text{ totalling } - 2.$$

Proceeding thus with each number we find 9 scores as follows:—

$$+3,\ -2,\ +5,\ -6,\ +3,\ 0,\ -1,\ +2,\ +1.$$

The total of these scores is + 5.

Now the maximum score obtained if the numbers are all in the objective order 1, 2, . . . 10, is 45. We therefore define the rank correlation coefficient τ as the ratio of the actual score to the maximum score, i.e., in the present case,

$$\tau = \frac{5}{45} = 0.111,$$

as compared with $\rho = 0{\cdot}139$ for the Spearman coefficient.

Generally, if there are n individuals the maximum score, obtained if and only if they

are in the order $(1, 2 \ldots n)$, is $(n-1) + (n-2) + \ldots + 1 = \frac{n(n-1)}{2}$. Denoting the actual score by S, we have then for the coefficient of rank correlation

$$\tau = \frac{2S}{n(n-1)} \quad . \quad (16.9)$$

16.7. The actual calculation of S may be shortened considerably. Looking again at the ranking (16.8) we see that the number 1 has two numbers on its right and seven on its left. We therefore score $2 - 7 = -5$ and strike out the 1. In the remaining ranking, the number 2 has 6 numbers on its right and two on its left, and hence we score $6 - 2 = +4$; we then strike out the 2 and proceed with the 3 as before. It will be found that the scores obtained are

$$-5, \ +4, \ +1, \ +6, \ -3, \ 0, \ +3, \ 0, \ -1.$$

The total of these scores is $+5$, and is equal to S. The rule is quite general. Its validity is evident from the consideration that instead of taking each number with its succeeding numbers we consider pairs contributing to S in a different way. Taking the number 1 first, and remembering that all other numbers are greater than 1, we see that any number on the left must contribute -1, and any number on the right $+1$, to S. When 1 is struck out the procedure remains valid for 2, and so on.

Alternatively the following procedure may be adopted. Considering again (16.8), we see that the number 4 has on its right 6 greater numbers, the 7 has 3 greater numbers, and so on, the numbers being

$$6, \ 3, \ 6, \ 0, \ 4, \ 2, \ 1, \ 2, \ 1,$$

totalling 25. There must therefore be $45 - 25 = 20$ numbers lying to the right of successive numbers in the ranking which are less than those numbers, and hence $S = 25 - 20 = 5$ as before. Generally, if the number obtained by counting greater numbers is k,

$$S = 2k - \frac{n(n-1)}{2}$$

and thus

$$\tau = \frac{4k}{n(n-1)} \quad . \quad (16.10)$$

A check may be obtained by counting greater numbers lying to the left. If the total of such numbers is l

$$S = \frac{n(n-1)}{2} - 2l$$

$$\tau = 1 - \frac{4l}{n(n-1)} \quad . \quad (16.11)$$

16.8. The extension of the use of τ to the case where no objective order is given requires a little further consideration. Suppose we have two rankings as follows :—

$$\left.\begin{array}{cccccccccccc} & A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & \ldots \\ P & 6 & 9 & 4 & 3 & 5 & 10 & 2 & 1 & 8 & 7 \\ Q & 6 & 5 & 10 & 2 & 3 & 9 & 7 & 4 & 1 & \end{array}\right\} \quad . \quad . \quad (16.12)$$

τ may be obtained by arranging one ranking in the natural order (1, 2 ... n) thus:

$$\left.\begin{array}{cccccccccccc} & A_8 & A_7 & A_4 & A_3 & A_5 & A_1 & A_{10} & A_9 & A_2 & A_6 \\ P' & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ Q' & 4 & 7 & 2 & 10 & 3 & 6 & 8 & 1 & 5 & 9 \end{array}\right\} \qquad (16.13)$$

and then finding τ between P' and Q' as in the preceding section. We have however to show that if we arrange Q in the natural order, giving

$$\left.\begin{array}{cccccccccccc} & A_9 & A_4 & A_5 & A & A_2 & A_1 & A_7 & A_{10} & A_6 & A_3 \\ P'' & 8 & 3 & 5 & 1 & 9 & 6 & 2 & 7 & 10 & 4 \\ Q'' & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array}\right\} \qquad (16.14)$$

then τ between P'' and Q'' is the same as that between P' and Q'. That this must be so may be seen as follows:—

In (16.13) the successive contributions to S are, as found by the method of **16.6**,

$$+3, \ -2, \ +5, \ -6, \ +3, \ 0, \ 1, \ +2, \ +1.$$

Consider now the contributions to S from (16.14) when the short method of **16.7** is used. They will be found to be exactly the same. If the permutation Q' begins with a_0 the contribution to $S_{Q'}$ from pairs involving a_0 will be $(n - a_0) - (a_0 - 1)$. In P'' the a_0th number will be 1 and the contribution to $S_{P''}$ will also be $(n - a_0) - (a_0 - 1)$. If the second number in Q' is a_1 the contribution to $S_{Q''}$ will be $(n - a_1) - (a_1 - 1) \pm 1$ according to whether a_1 is greater than a_0 or not. In P'' the a_1th number will be 2 and the contribution to $S_{P''}$ is also $(n - a_1) - (a_1 - 1) \pm 1$ according to whether 1 lies on the left or the right of 2 in P'', i.e. whether a_1 is greater than a_0 or not; and so on.

In practical calculations it is not necessary to carry out the rearrangements. Consider again (16.12). The number 1 in Q has an 8 above it in P. In the ranking of the A's 8 has two members to the right and seven to the left. Score therefore, -5, and strike out A_8. The number 2 in Q has a 3 above it in P, and A_3 has six members to its right (ignoring A_8) and two to its left, score $+4$; and so on, the scores being

$$-5, \ +4, \ +1, \ +6, \ -3, \ 0, \ +3, \ 0, \ -1$$

totalling $+5$ which is equal to S.

16.9. Like ρ, τ is $+1$ only if the correspondence between two rankings is perfect and -1 only if the rankings are inverted. In actual practice the values given by the two coefficients bear a nearly constant ratio (cf. **16.24**) and one appears to be as good as the other so far as providing a measure of ranking concordance is concerned. ρ is, however, easier to calculate and is probably the most convenient to use. Against this must be set certain difficulties in its sampling distribution, which will be referred to below, and the fact that τ can be generalised to the case of partial rank correlations.

16.10. In considering the interpretation of any particular value of ρ or τ the question naturally arises, are such values significant in the statistical sense, i.e. can they have arisen by chance from a population in which the qualities under consideration are independent? And further, can we assign a standard error to the observed values? The second question is not an easy one to answer, or even to understand unless ranks are related to variate-values. In the sampling of variates we are given a set of n values emanating from a population of values. In the ranking case we are given n ordinal numbers, but it is useless

to consider them as emanating from a population of (different) ordinal numbers. The point will be considered later when we introduce the concept of grades (**16.25**).

The sampling problem, however, acquires a definite meaning if the two qualities under consideration are independent. In such a case the pairs of rankings of n members drawn at random are independent; and consequently in a large number of samples there will occur in equal amounts every ranking according to one quality associated with every ranking according to the other. We are thus led to consider the distributions of ρ and τ in populations consisting of all possible associations of all possible rankings. Clearly no generality is lost if we fix one ranking as the order $(1, 2 \ldots n)$ and consider its correlations with the $n!$ possible permutations of those numbers. If a given ρ or τ cannot, to an acceptable degree of probability, have arisen from such a population, we are justified in concluding that the two qualities have some definite relationship in the population.

Sampling Distribution of Spearman's ρ in the Case of Independence

16.11. Consider then the distribution of values of ρ in the population obtained by correlating the order $(1, 2 \ldots n)$ with every possible permutation of the n natural numbers. We shall, in fact, find it more convenient to consider the distribution of $\Sigma(d^2)$, which is simply related to ρ by equation (16.5). Certain elementary properties of the distribution are obtainable immediately.

(*a*) Any value of $\Sigma(d^2)$ must be even; for $\Sigma(d) = 0$ and hence the number of odd values of d, and thus of d^2, is even.

(*b*) The possible values of $\Sigma(d^2)$ range from 0 to $\frac{1}{3}(n^3 - n)$ and hence there are $\frac{1}{6}(n^3 - n) + 1$ of them.

(*c*) The distribution is symmetrical, about a central value if $\frac{1}{6}(n^3 - n)$ is even, or about two adjacent central values if it is odd. This follows from the fact that to any value of ρ corresponding to a permutation P there will correspond a negative value of ρ, of the same absolute value, arising from P inverted. For if P is $X_1\, X_2 \ldots X_n$, the inverted permutation is $X_n, X_{n-1}, \ldots X_1$. $\Sigma(d^2)$ calculated from P is then $\sum_{i=1}^{n}(X_i - i)^2$ and that from ρ inverted is $\sum_{i=1}^{n}(X_i - \overline{n+1} + i)^2$. The sum of these two is

$$\Sigma(X_i^2) + \Sigma(i^2) - 2\Sigma(X_i i) + \Sigma(X_i^2) + \Sigma(n + 1 - i)^2 - 2\Sigma\left\{X_i(n + 1 - i)\right\}.$$

The first, second, fourth and fifth terms in this expression are equal to $\Sigma(i^2)$, i.e. to $\frac{1}{6}n(n + 1)(2n + 1)$. The sum of the third and sixth is

$$- 2(n + 1)\,\Sigma(X) = - n(n + 1)^2.$$

Thus the sum of the two $\Sigma(d^2)$ is

$$\begin{aligned}&\tfrac{2}{3}n(n + 1)(2n + 1) - n(n + 1)^2\\ &= \tfrac{1}{3}(n^3 - n).\end{aligned}$$

Thus we see from (16.5) that the sum of the corresponding ρ's is zero.

(*d*) It follows that all odd moments of the distribution of $\Sigma(d^2)$ about the mean vanish.

16.12. Consider the deviations between the order $1, 2, \ldots, n$ and an order X. If one deviation is known, then certain deviations become impossible for other ranks. For instance, if the deviation d_1 between X_1 and 1 is $(n - 1)$, then $X_1 = n$, and it is impossible

for the deviation between X_2 and 2 to be $(n-2)$; or for the deviation between X_3 and 3 to be $(n-3)$, and so on. Consider then the array:

$$\begin{array}{ccccccc}
n-1 & n-2 & n-3 \dots & 2 & 1 & 0 \\
n-2 & n-3 & n-4 \dots & 1 & 0 & -1 \\
n-3 & n-4 & n-5 \dots & 0 & & -2 \\
 & & & & & \\
2 & 1 & 0 \quad . & -(n-5) & -(n-4) & -(n-3) \\
1 & 0 & -1 \quad . & -(n-4) & -(n-3) & -(n-2) \\
0 & -1 & -2 \quad . & -(n-3) & -(n-2) & -(n-1)
\end{array}$$

If d_k has the value in the rth row and the kth column, then d_l cannot have the value in the rth row and the lth column; and so on.

In fact, any permissible set of deviations is given by taking n entries from the above table so that no row or column contributes more than one entry.

Hence to get $\Sigma(d^2)$ for any permissible set, write

$$E = \begin{array}{cccc}
a^0 & & & a^{(n-1)^2} \\
 & a^1 & & .\; a^{(n-2)^2} \\
 & & & a^{(n-3)^2} \\
 & & & \\
a^{(n-1)^2} & a^{(n-2)^2} & a^{(n-3)^2} & a^{(n-4)^2}
\end{array}$$

and $\Sigma(d^2)$ is given by the index of a of one of the terms obtained from E by choosing n factors so that no row or column appears more than once and multiplying them together. Thus the distribution of $\Sigma(d^2)$ is given by the totality of $n!$ terms which can be constructed in that way. E will be taken to be equal to the polynomial in a given by the sum of these terms—the so-called "permanent."

16.13. E bears an obvious analogy to the determinant, but it cannot be regarded as such and expanded accordingly. If it could, the distribution of $\Sigma(d^2)$ would be obtained without difficulty, for a determinant with the elements of E as given above may be shown to be equal to

$$(1-a^2)^{n-1}(1-a^4)^{n-2}(1-a^6)^{n-4}\dots(1-a^{2(n-1)}).$$

E, in fact, lacks the fundamental property of the determinant in that it does not change sign if two rows or columns are interchanged.

Nevertheless certain of the rules of determinantal algebra remain true for E. The most valuable is that E may be expanded in terms of its minors of any order in the usual way. Expansion of this type is, in fact, rather easier with E than with the determinant, for all terms of E are essentially positive and there are no difficulties with signs. Such expansions were used in obtaining the distributions given below. There are also certain devices which assist the expansion of E in virtue of its symmetry. Two which will be found useful are as follows:—

(*a*) Any minor of E is symmetrical in powers of a, i.e. is of the form

$$A_0a^k + A_2a^{k-2} + A_4a^{k-4} + \dots + A_4a^{m-4} + A_2a^{m-2} + A_0a^m.$$

(*b*) The effect of shifting a minor bodily across E is to multiply each term of its expansion by a constant power of a,

TABLE 16.1

Spearman's ρ. Distribution of $\Sigma(d^2)$ for Values of n from 1 to 8.

Values of n.

$\Sigma(d^2)$	1	2	3	4	5	6	7	8
0	1	1	1	1	1	1	1	1
2	.	1	2	3	4	5	6	7
4	.	.	0	1	3	6	10	15
6	.	.	2	4	6	9	14	22
8	.	.	1	2	7	16	29	47
10	.	.	.	2	6	12	26	54
12	.	.	.	2	4	14	35	70
14	.	.	.	4	10	24	46	94
16	.	.	.	1	6	20	55	129
18	.	.	.	3	10	21	54	124
20	.	.	.	1	6	23	74	178
22	.	.	.	.	10	28	70	183
24	.	.	.	.	6	24	84	237
26	.	.	.	.	10	34	90	238
28	.	.	.	.	4	20	78	276
30	.	.	.	.	6	32	90	264
32	.	.	.	.	7	42	129	379
34	.	.	.	.	6	29	106	349
36	.	.	.	.	3	29	123	380
38	.	.	.	.	4	42	134	400
40	.	.	.	.	1	32	147	517
42	.	.	.	.	.	20	98	394
44	.	.	.	.	.	34	168	542
46	.	.	.	.	.	24	130	492
48	.	.	.	.	.	28	175	640
50	.	.	.	.	.	23	144	557
52	.	.	.	.	.	21	168	666
54	.	.	.	.	.	20	144	595
56	.	.	.	.	.	24	184 (median)	776
58	.	.	.	.	.	14	.	684
60	.	.	.	.	.	12	.	786
62	.	.	.	.	.	16	.	718
64	.	.	.	.	.	9	.	922
66	.	.	.	.	.	6	.	745
68	.	.	.	.	.	5	.	917
70	.	.	.	.	.	1	.	781
72	.	.	.	.	.	.	.	982
74	.	.	.	.	.	.	.	826
76	.	.	.	.	.	.	.	950
78	.	.	.	.	.	.	.	844
80	.	.	.	.	.	.	.	1066
82	.	.	.	.	.	.	.	845
84	.	.	.	.	.	.	.	936 (median)
TOTALS	1	2	6	24	120	720	5040*	40,320*

* Total of whole distribution, only the median value and the values on one side of the median being shown in this table.

TABLE 16.2

Spearman's ρ. Probability that $\Sigma(d^2)$ will be Attained or Exceeded for Values of n from 4 to 8 inclusive.

$\Sigma(d^2)$.

	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28
$n = 4$	1	0·958	0·833	0·792	0·625	0·542	0·458	0·375	0·208	0·167	0·042				
$n = 5$	1	0·992	0·958	0·933	0·883	0·825	0·775	0·742	0·658	0·608	0·525	0·475	0·392	0·342	0·258
$n = 6$	1	0·999	0·992	0·983	0·971	0·949	0·932	0·912	0·879	0·851	0·822	0·790	0·751	0·718	0·671
$n = 7$	1	1·000	0·999	0·997	0·994	0·988	0·983	0·976	0·967	0·956	0·945	0·931	0·917	0·900	0·882
$n = 8$	1	1·000	1·000	0·999	0·999	0·998	0·996	0·995	0·992	0·989	0·986	0·982	0·977	0·971	0·965

	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58
$n = 5$	0·225	0·175	0·117	0·067	0·042	$0{\cdot}0^{2}83$									
$n = 6$	0·643	0·599	0·540	0·500	0·460	0·401	0·357	0·329	0·282	0·249	0·210	0·178	0·149	0·121	0·088
$n = 7$	0·867	0·849	0·823	0·802	0·778	0·751	0·722	0·703	0·669	0·643	0·609	0·580	0·547	0·518	0·482
$n = 8$	0·958	0·952	0·943	0·934	0·924	0·915	0·902	0·892	0·878	0·866	0·850	0·837	0·820	0·805	0·786

	60	62	64	66	68	70	72	74	76	78	80	82	84	86	88
$n = 6$	0·068	0·051	0·029	0·017	$0{\cdot}0^{2}83$	$0{\cdot}0^{2}14$									
$n = 7$	0·453	0·420	0·391	0·357	0·331	0·297	0·278	0·249	0·222	0·198	0·177	0·151	0·133	0·118	0·100
$n = 8$	0·769	0·750	0·732	0·709	0·690	0·668	0·648	0·624	0·603	0·580	0·559	0·533	0·512	0·488	0·467

	90	92	94	96	98	100	102	104	106	108	110	112	114	116	118
$n = 7$	0·083	0·069	0·055	0·044	0·033	0·024	0·017	0·012	$0{\cdot}0^{2}62$	$0{\cdot}0^{2}34$	$0{\cdot}0^{2}14$	$0{\cdot}0^{3}20$			
$n = 8$	0·441	0·420	0·397	0·376	0·352	0·332	0·310	0·291	0·268	0·250	0·231	0·214	0·195	0·180	0·163

	120	122	124	126	128	130	132	134	136	138	140	142	144	146	148
$n = 8$	0·150	0·134	0·122	0·108	0·098	0·085	0·076	0·066	0·057	0·048	0·042	0·035	0·029	0·023	0·018

	150	152	154	156	158	160	162	164	166	168
$n = 8$	0·014	0·011	$0{\cdot}0^{2}77$	$0{\cdot}0^{2}54$	$0{\cdot}0^{2}36$	$0{\cdot}0^{2}23$	$0{\cdot}0^{2}11$	$0{\cdot}0^{3}57$	$0{\cdot}0^{3}20$	$0{\cdot}0^{4}25$

e.g. the minors

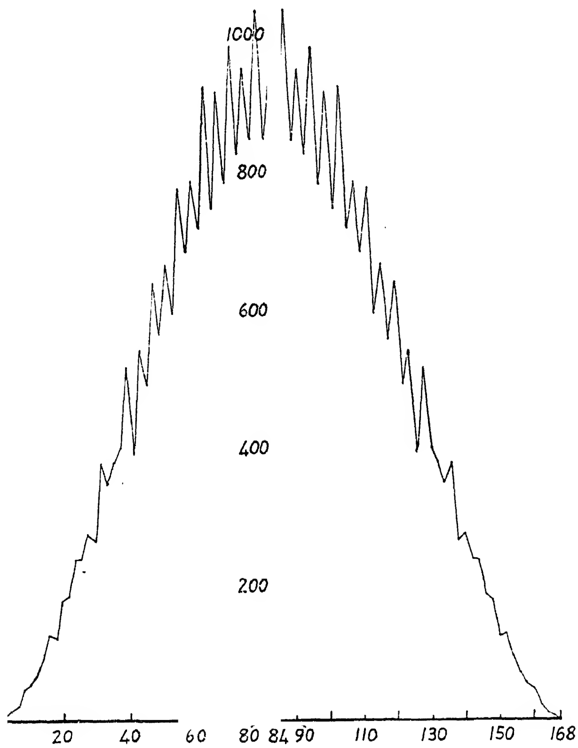
$$M = \qquad\qquad = a^0 + 2a^2 + 2a^4 + a^6$$

and
$$M' = \quad a^1 \qquad\qquad = a^{12}(a^0 + 2a^2 + 2a^4 + a^6)$$

are related by
$$M' = Ma^{12}.$$

16.14. The tables on pp. 396–7 show the frequencies of $\Sigma(d^2)$ for values of n from 1 to 8 inclusive and the probabilities that a given value of $\Sigma(d^2)$ will be attained or exceeded on random sampling for n from 4 to 8 inclusive.

16.15. The distributions of Table 16.1 are peculiar in several respects. For lower values of n they are distinctly bimodal. For $n = 7$ and $n = 8$ the frequency polygons have an unusual serrated profile, that for the latter being shown in Fig. 16.1, though normality



Values of $\Sigma(d^2)$.

FIG. 16.1. Spearman's ρ. Frequency Polygon of $\Sigma(d^2)$ for $n = 8$.

is beginning to emerge. It will be shown below that as $n \to \infty$ the distribution tends to normality, but it is not immediately obvious how a serrated polygon of this kind can do so.

I think that the tails of the curve smooth out first, and that as n increases the smoothness runs up the curve towards the apex.

16.16. The calculation of frequencies for n greater than 8 would be a tedious process and can be obviated by finding curves which satisfactorily approximate to the distribution, at least so far as its distribution function is concerned. For this purpose we will find the second and fourth moments of ρ about its mean. The first and third, of course, are zero.

Suppose we measure the rank numbers from their mean, writing for the new variables $x = X - \frac{1}{2}(n + 1)$, $y = Y - \frac{1}{2}(n + 1)$. Then from **16.4** we have

$$= \frac{12\Sigma(xy)}{n^3 - n} = \frac{1}{N}\Sigma(xy), \text{ say,}$$

where $N = \dfrac{(n^3 - n)}{12}$. Since $E(\rho) = \mu'_1(\rho) = 0$ we have

$$\operatorname{var}\rho = \frac{1}{N^2}E(\Sigma xy)^2$$

$$= \frac{1}{N^2}E\{\Sigma(x^2y^2)\} + \frac{1}{N^2}E\{\Sigma(x_ix_j\ y_iy_j)\} \quad . \quad (16.15)$$

where $i \neq j$. Now for any value of x, y may have any value from 1 to n. Hence

$$E\Sigma(x^2y^2) = nE(x^2)E(y^2)$$

$$= \frac{1}{n}\{\Sigma(x^2)\}^2$$

$$= \frac{N^2}{n} \quad . \quad (16.16)$$

Further, in the product term of (16.15) there are $n(n - 1)$ pairs of values $i \neq j$ and thus

$$E\Sigma(x_ix_j\ y_iy_j) = n(n-1)\ E(x_ix_j\ y_iy_j)$$

$$= n(n-1)\ E(x_ix_j)^2$$

$$\frac{1}{n(n-1)}\ (\Sigma x_ix_j)^2$$

$$\frac{1}{n(n-1)}\{(\Sigma x)^2 - \Sigma(x^2)\}^2$$

$$- \frac{N^2}{n(n-1)} \quad . \quad . \quad . \quad (16.17)$$

Hence, substituting from (16.16) and (16.17) in (16.15) we have

$$\operatorname{var}\rho = \frac{1}{n} + \frac{1}{n(n-1)}$$

$$\frac{1}{n-1} \quad (16.18)$$

By the same technique it may be shown that

$$\mu_4(\rho) = \frac{3(25n^3 - 38n^2 - 35n \quad 72)}{25n(n+1)(n-1)^3} \quad (16.19)$$

16.17. Consider now the Type II symmetric distribution

$$dF = \frac{1}{B\left(\frac{1}{2}, \frac{n-2}{2}\right)} (1 - x^2)^{\frac{n-4}{2}}\, dx, \qquad -1 \leqslant x \leqslant 1 \tag{16.20}$$

The first and third moments are, of course, zero. The second and fourth are given by

$$\mu_2 = \frac{B\left(\frac{3}{2}, \frac{n-2}{2}\right)}{B\left(\frac{1}{2}, \frac{n-2}{2}\right)} = \frac{1}{n-1} \tag{16.21}$$

$$\mu_4 = \frac{B\left(\frac{5}{2}, \frac{n-2}{2}\right)}{B\left(\frac{1}{2}, \frac{n-2}{2}\right)} = \frac{3}{n^2-1} \tag{16.22}$$

The distribution thus has its first three moments the same as those of Spearman's ρ in the case of independence. The fourth moments are the same to order n^{-2}, the difference being

$$\frac{3}{n^2-1}\left\{1 - \frac{25n^3 - 38n^2 - 35n + 72}{25n(n-1)^2}\right\} \quad -36$$

i.e. of lower order in n than the moments themselves. It has therefore been suggested that the distribution (16.20) may be used instead of that of ρ to give the distribution function of the latter for moderate or large n. Tests on the distributions of Table 16.1 indicate that this is a justifiable approximation.

For instance, when $n = 8$ the distribution (16.20) becomes

$$dF = \frac{1}{B(\frac{1}{2}, 3)} (1 - x^2)^2\, dx$$

and by direct integration the probability of obtaining a value of x greater than x_0 in absolute value is

$$1 - \frac{15}{8}\left(x_0 - \frac{2x_0^3}{3} + \frac{x_0^5}{5}\right) \tag{16.23}$$

In comparing this with the values of the ρ-distribution it is as well to make a continuity correction, similar to that of **12.15**, to allow for the fact that the distribution of ρ is discontinuous whereas that of x is continuous. If the values of $\Sigma(d^2)$ are regarded as spread over a range of one unit on each side of the actual value, the range of $\Sigma(d^2)$ is increased from $\frac{1}{3}(n^3 - n)$ to $\frac{1}{3}(n^3 - n) + 2$, each terminal contributing a unit. Instead of writing $x = \rho$ we will then write

$$x = 1 - \frac{\Sigma(d^2)}{\frac{1}{6}(n^3 - n) + 1}. \qquad . \quad . \quad . \quad . \quad (16.24)$$

Now from Table 16.2 the probability of obtaining a value of ρ greater than $\frac{1}{6}$ in absolute value, corresponding to $\Sigma(d^2)$ outside the range 14 to 154 inclusive, is $2 \times 0{\cdot}0053 = 0{\cdot}0106$. The appropriate x from (16.24) is $1 - \frac{14}{85} = 0{\cdot}835$, and this on substitution in (16.23) gives the probability of 0·0098. Similarly the chance of getting a value of $\Sigma(d^2)$ outside the range 26 to 142 inclusive is 0·0576. That given by (16.23) is 0·0561. The agreement is evidently good enough for most practical purposes and would, of course, improve as n increases.

16.18. If we put, in (16.20),

$$= x\left(\frac{n-2}{1-x^2}\right)^{\frac{1}{2}}$$

we obtain the distribution

$$dF = \frac{1}{(n-2)^{\frac{1}{2}}\, B(\frac{1}{2}, \frac{1}{2}n - 1)} \frac{dt}{\left(1 + \frac{t^2}{n-2}\right)^{\frac{n-1}{2}}}, \quad . \quad . \quad . \quad (16.25)$$

the "Student" distribution of Example 10.6. If n is large the continuity correction may be neglected and to this approximation

$$x = \rho,$$

so that ρ may be tested in "Student's" distribution by writing

$$t = \rho\left(\frac{n-2}{1-\rho^2}\right)^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16.26)$$

Example 16.3

In Example 16.2 we found a value of $\rho = 0{\cdot}139$. Is this significant ?

We have $n = 10$ and from (16.26)

$$t = 0{\cdot}139\sqrt{\frac{8}{1-(0{\cdot}139)^2}}$$

$$= 0{\cdot}397.$$

From Appendix Table 3 we see that the chance of getting such a value or greater in absolute value is about $0{\cdot}70$ $(= 2(1 - 0{\cdot}65))$. The value cannot therefore be regarded as significant.

16.19. As n tends to infinity the B-distribution tends to the normal form and we therefore suspect that ρ also tends to normality. That this is in fact so may be seen as follows ; the proof being due to Hotelling and Pabst (1936).

The general moment of ρ of even order is given by

$$\mu_{2\alpha} = \frac{1}{S_2^{2\alpha}} E(x_1 y_1 + \ . \quad . \ x_n y_n)^{2\alpha} \quad . \quad (16.27)$$

where S_2 is written for $\sum_{i=1}^{n} x_i^2$ and generally S_p for $\sum_{i=1}^{n} x_i^p$. When the parenthesis is expanded we may, in virtue of the independence of x and y, take expectations term by term, regarding the x's as constant. Now

$$E(y_i^{2\alpha}) = E(x_i^{2\alpha}) = \frac{1}{n}\Sigma(y_i^{2\alpha}) = \frac{1}{n}S_{2\alpha}$$

$$E(y_i^{2\alpha-1}\, y_j) = \frac{1}{n(n-1)}\Sigma(y_i^{2\alpha}\, y_j), \text{ etc.}$$

Hence

$$\mu_{2\alpha} = \frac{1}{S_2{}^{2\alpha}}\left\{\frac{1}{n}(\Sigma x_i{}^{2\alpha})^2 + \frac{A}{n(n-1)}(\Sigma x_i{}^{2\alpha-1}x_j)^2 + \frac{B}{n(n-1)(n-2)}(\Sigma x_i{}^{2\alpha-2}x_j x_k)^2 + \text{etc.}\right\} \qquad (16.28)$$

where the coefficients A depend on α but not on n. We proceed to show that the term of greatest degree in n in (16.28) is the term $\Sigma(x_i{}^2 x_j{}^2 \ldots x_p{}^2)$.

The numerator of any term in (16.28), being a symmetric function of the x's, can be expressed in terms of the symmetric sums S_p. Further S_p vanishes if p is odd. Since any S_k is of degree $k+1$ in n, the degree of a non-vanishing term $S_{\alpha_1} S_{\alpha_2} \ldots S_{\alpha_p}$ is $\sum_{j=1}^{p}(\alpha_j + 1) = 2\alpha + p$. Consequently the term of highest degree in n must contain as high a p as possible, that is to say as many S's as possible, subject to the requirement that the subscript of each S must be even.

Now consider a term

$$\Sigma(x_1{}^{\alpha_1} \ldots x_p{}^{\alpha_p}) = \Sigma c_0 S_{\alpha_1} S_{\alpha_2} \ldots S_{\alpha_p} + \Sigma c_1 S_{\alpha_1+\alpha_2} S_{\alpha_3} \ldots S_{\alpha_p} + \Sigma c_2 S_{\alpha_1+\alpha_2+\alpha_3} S_{\alpha_4} \ldots S_{\alpha_p} + \ldots, \text{ etc.} \qquad (16.29)$$

If the α's are all even the term of highest degree on the right is, as just remarked, $2\alpha + p$. If the α's are not all even, suppose there are m even ones and $2q$ odd ones $(m + 2q = p)$. Then the first term in (16.29) vanishes and the term of highest degree which does not vanish must be obtained by grouping q pairs of odd α's, and hence is of degree $2\alpha + m + q = 2\alpha + p - q$.

Now in (16.28) the degree of the denominator in each term is the number of different x's in the numerator. Thus the term of highest degree in x is of degree

$$\begin{aligned} 2(2\alpha + p - q) - (m + 2q) &= 4\alpha - m + 2p - 4q \\ &= 4\alpha + m. \end{aligned}$$

This will be a maximum when m is a maximum and therefore when q is zero, in which case $m = \alpha$. Hence the greatest degree in n in (16.28) arises from the term $\Sigma(x_i{}^2 x_j{}^2 \ldots x_p{}^2)$ as stated. Now in the expansion of

$$(x_1 y_1 + \ldots x_n y_n)^2$$

the coefficient of $x_1^2 \quad x_\alpha{}^2 y_1^2 \ldots y_\alpha{}^2$ is, by the multinomial theorem, $\frac{(2\alpha)!}{2^\alpha}$ and hence

$$\mu_{2\alpha} \quad \frac{1}{S_2{}^{2\alpha}} \frac{(2\alpha)!}{2^\alpha} \{\Sigma(x_1^2 \ldots x_\alpha{}^2)\}^2 \qquad (16.30)$$

The term of highest degree in n in $\Sigma(x_1^2 \ldots x_\alpha{}^2)$ is that in $S_2{}^\alpha$, the coefficient of which is evidently the reciprocal of that of $\Sigma(x_1^2 \ldots x_\alpha{}^2)$ in

$$S_2{}^\alpha = (x_1^2 + \ldots x_n{}^2)^\alpha,$$

i.e.

$$= \alpha!$$

Thus, from (16.30),

$$\mu_{2\alpha} \sim \frac{(2\alpha)!}{2^\alpha . \alpha!}\left\{\frac{1}{n^\alpha} + O\left(\frac{1}{n^{\alpha+1}}\right)\right\}$$

Now $\mu_2 = \dfrac{1}{n-1}$ and thus

$$\frac{\mu_{2\alpha}}{\mu_2{}^{\alpha}} \to \frac{(2\alpha)!}{2^{\alpha}\alpha!} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16.31)$$

i.e. to the moments of the normal distribution of unit variance. It follows from the Second Limit Theorem of **4.24** that the distribution of ρ tends to normality. The tendency is not, however, very rapid and we have already noticed the peculiar character of the distribution for lower n.

Distribution of τ in the Case of Independence

16.20. We now consider the distribution of the coefficient τ under similar conditions, that is to say in a population obtained by correlating a given ranking with all the $n!$ possible rankings.

Consider a given ranking of the numbers 1, 2, . . . n and the effect of inserting an additional number $(n + 1)$ in the various possible places in the ranking, from the first place (preceding the first number) to the last place (following the last number).

Inserting a number at the beginning will add $-n$ to the value of S of equation (16.9). Inserting it between the first and second will add $-(n - 2)$ to S; and so on. Thus to any frequency-distribution of S for given n, say $f(S, n)$, there will correspond frequencies $f(S - n, n), f(S - (n - 2), n) \ . \ . \ . \ f(S + n, n)$, the sum of which gives $f(S, n + 1)$. If the frequency of a given S is the coefficient of x^S in a polynomial $P(x)$, then the corresponding values of S in the frequency for $(n + 1)$ are the coefficients of

$$(x^{-n} + x^{-(n-2)} + \ . \ . \ . + x^{n-2} + x^{n})P(x).$$

But the frequency-distribution of S when $n = 2$ is given by $x^{-1} + x^{1}$, there being one value $S = -1$ and one value $S = 1$. Thus the frequencies of S for rankings of n are the coefficients of x^S in the array

$$\begin{aligned} f = (x^{-1} + x)(x^{-2} + 1 + x^{2})(x^{-3} + x^{-1} + x^{1} + x^{3}) \ . \ . \ . \ (x^{-(n-1)} + x^{-(n-3)} \\ + \ . \ . \ . + x^{(n-3)} + x^{(n-1)}) \end{aligned} \quad . \quad . \quad . \quad (16.32)$$

It follows that the distribution of S, and hence that of τ, is symmetrical about zero. The values of S are either all odd or all even, according to whether $\dfrac{n(n-1)}{2}$ is odd or even. The actual frequencies may be calculated by a figurate triangle, as follows :—

Value of n	Frequencies of S										
1	1										
2	1	1									
3	1	2	2	1							
4	1	3	5	6	5	3	1				
	1	4	9	15	20	22	20	15	9	4	1

(16.33)

In this array a number in the rth row is the sum of the r numbers above it and the $(r - 1)$ numbers to the left of that number. A little reflection will show that this rule follows from (16.32). The formation of the array is quite simple and several devices shorten the arithmetic. For instance, in part of the array towards the left a number in the rth row is the sum of the number immediately above it and the number immediately to the left. The array is symmetrical and the total in the rth row is $r!$

The following tables show the frequency-distribution of S for values of n from 1 to 10 inclusive and the probability that a value of S will be attained or exceeded.

TABLE 16.3

Rank Coefficient τ. Distribution of S for Values of n from 1 to 10 (only the Positive Half of the Symmetrical Distribution shown).

S	Values of n					S	Values of n				
	1	4	5	8	9		2	3	6	7	10
0	1	6	22	3,836	29,228	1	1	2	101	573	250,749
2		5	20	3,736	28,675	3		1	90	531	243,694
4		3	15	3,450	27,073	5			71	455	230,131
6		1	9	3,017	24,584	7			49	359	211,089
8			4	2,493	21,450	9			29	259	187,959
10			1	1,940	17,957	11			14	169	162,337
12				1,415	14,395	13			5	98	135,853
14				961	11,021	15			1	49	110,010
16				602	8,031	17				20	86,054
18				343	5,545	19				6	64,889
20				174	3,606	21				1	47,043
22				76	2,191	23					32,683
24				27	1,230	25					21,670
26				7	628	27					13,640
28				1	285	29					8,095
30					111	31					4,489
32					35	33					2,298
34					8	35					1,068
36					1	37					440
						39					155
						41					44
						43					9
						45					1

16.21. As may be seen by comparing Tables 16.1 and 16.3, the distribution of S, and hence that of τ, is much smoother than that of $\Sigma(d^2)$ and ρ. We show below that it tends to normality, and in fact the tendency is so rapid that for values of n greater than 10 the normal distribution provides an adequate approximation. We proceed to find the second and fourth moment of the distribution.

If we differentiate the expression f in (16.32) and equate x to 1 we evidently obtain the first moment of S; and generally, writing θ for the operator $x\dfrac{\partial}{\partial x}$,

$$n!\mu_r = (\theta^r f)_{x=1} \tag{16.34}$$

For example, when $r = 1$ we have

$$\begin{aligned} n!\,\mu_1 &= (-1+1)(1+1+1) \qquad (1+1+\ldots+1) \\ &\quad + (1+1)(-2+2)(\;.\;. \\ &\qquad + \text{etc.} \\ &= 0. \end{aligned}$$

TABLE 16.4

Probability that S attains or exceeds a Specified Value. (Shown only for Positive Values. Negative Values obtainable by Symmetry.)

S	Values of n: 4	5	8	9	S	Values of n: 6	7	10
0	0·625	0·592	0·548	0·540	1	0·500	0·500	0·500
2	0·375	0·408	0·452	0·460	3	0·360	0·386	0·431
4	0·167	0·242	0·360	0·381	5	0·235	0·281	0·364
6	0·042	0·117	0·274	0·306	7	0·136	0·191	0·300
8		0·042	0·199	0·238	9	0·068	0·119	0·242
10		$0{\cdot}0^{2}83$	0·138	0·179	11	0·028	0·068	0·190
12			0·089	0·130	13	$0{\cdot}0^{2}83$	0·035	0·146
14			0·054	0·090	15	$0{\cdot}0^{2}14$	0·015	0·108
16			0·031	0·060	17		$0{\cdot}0^{2}54$	0·078
18			0·016	0·038	19		$0{\cdot}0^{2}14$	0·054
20			$0{\cdot}0^{2}71$	0·022	21		$0{\cdot}0^{3}20$	0·036
22			$0{\cdot}0^{2}28$	0·012	23			0·023
24			$0{\cdot}0^{3}87$	$0{\cdot}0^{2}63$	25			$0{\cdot}0^{2}14$
26			$0{\cdot}0^{3}19$	$0{\cdot}0^{2}29$	27			$0{\cdot}0^{2}83$
28			$0{\cdot}0^{4}25$	$0{\cdot}0^{2}12$	29			$0{\cdot}0^{2}46$
30				$0{\cdot}0^{3}43$	31			$0{\cdot}0^{2}23$
32				$0{\cdot}0^{3}12$	33			$0{\cdot}0^{2}11$
34				$0{\cdot}0^{4}25$	35			$0{\cdot}0^{3}47$
36				$0{\cdot}0^{5}28$	37			$0{\cdot}0^{3}18$
					39			$0{\cdot}0^{4}58$
					41			$0{\cdot}0^{4}15$
					43			$0{\cdot}0^{5}28$
					45			$0{\cdot}0^{6}28$

When $r = 2$ the operation on f will result in two types of terms, those in which both operations operate on one factor of f and those in which the operations operate on separate factors. When $x = 1$ these last vanish and thus

$$n!\,\mu_2 = (1+1)\frac{n!}{2} + (2^2+2^2)\frac{n!}{3} + \frac{(3^2+1^2+1^2+3^2)n!}{4}$$
$$+ \dots \; (\overline{n-1}^2 + \overline{n-3}^2 + \dots + \overline{n-3}^2 + \overline{n-1}^2)n!$$

$$\mu_2 = \frac{2}{2}.1^2 + \frac{2}{3}.2^2 + \frac{2}{4}(1^2+3^2) + \dots + \frac{2}{n}(\overline{n-1}^2 + \overline{n-3}^2 + \dots)$$

This may be summed by the ordinary methods of elementary algebra, and we find

$$\mu_2 = \frac{n(n-1)(2n+5)}{18} \qquad . \quad . \quad . \quad . \quad . \quad (16.35)$$

In a like manner it appears that

$$\mu_4 = \frac{n(n-1)}{2}\left\{1 + \frac{74}{9}(n-2) + \frac{37}{6}(n-2)(n-3) + \frac{32}{25}(n-2)(n-3)(n-4)\right.$$
$$\left. + \frac{2}{27}(n-2)(n-3)(n-4)(n-5)\right\} \qquad . \quad . \quad . \quad (16.36)$$

16.22. To prove that the distribution of τ tends to normality as $n \to \infty$ we shall show that

$$\mu_{2\alpha} \to \frac{(2\alpha)!}{2^\alpha \alpha!}(\mu_2)^\alpha.$$

Consider the effect of operating on f in (16.32) by θ 2α times and then putting $x = 1$. There will appear terms like

$$n! \quad + (r-2)^{2\alpha} + \ldots + (r-2)^{2\alpha} + r^{2\alpha}\}$$

$$n! \quad \frac{-(r-2)^{2\alpha-1} \ldots + (r-2)^{2\alpha-1} + r^{2\alpha-1}}{r}\Big\}\Big\{\frac{-t-(t-2)+\ldots(t+2)+t}{t}$$

etc. Any term with an odd superscript vanishes. Consider now the sum of terms like

$$n! \{r^2 + \ldots + r^2\}\{t^2 + \ldots + t^2\} \quad \{u^2 + \ldots + u^2 \qquad (16.37)$$
$$t \qquad u$$

It will be shown below that this term contributes the greatest power of n to the sum giving $n!\,\mu_{2\alpha}$.

In virtue of the multinomial form of Leibniz' theorem on the differentiation of a product, the factor by which this term is multiplied in the expansion of $\theta^{2\alpha}f$ is

$$\frac{(2\alpha)!}{2!\ldots 2!} = \frac{(2\alpha)!}{2^\alpha}$$

Hence $$\mu_{2\alpha} \quad \frac{(2\alpha)!}{2^\alpha}\,\{\text{Sum of terms like (16.37)}\} \qquad (16.38)$$

Each of these terms is of type $\frac{1}{r}\{r^2 + (r-2)^2 + \ldots (r-2)^2 + r^2\}$ i.e. is of order $\frac{r^2}{3}$.

The sum will then tend to the sum of terms like $\frac{1}{3^\alpha}(1^2 . 2^2 \ldots \alpha^2)$, each term containing α squares of the numbers $1, 2 \ldots n-1$. Call this π_α.

Then π_α is $\frac{1}{\alpha!}$ times the sum of terms in

$$\frac{1}{3^\alpha}\{1^2 + 2^2 + \ldots n - 1^2\}^\alpha \qquad (16.39)$$

which contain α different factors.

Now (16.38) is of order $\frac{n^{3\alpha}}{9^\alpha} \sim \mu_2{}^\alpha$. Hence if π_a tends to equality with the sum (16.39)

$$\pi, \sim \frac{\mu_2{}^\alpha}{\alpha!}$$

and hence, from (16.38)

$$\mu_{2\alpha} \quad \frac{(2\alpha)!}{2^\alpha}\frac{(\mu_2)^\alpha}{\alpha!}.$$

We have then to show that (16.38) tends asymptotically to the sum of its terms $\alpha!\,\pi_\alpha$, i.e. that sums of terms like

$$1^4 . 2^2 \ldots (\alpha-1)^2, \quad 1^6 . 2^2 \ldots (\alpha-2)^2$$

tend in comparison to zero. This may be shown inductively. Consider first of all

$$\{1^2 + 2^2 + \ldots (n-1)^2\}^2 = 2\pi_2 + 1^4 + 2^4 + \ldots (n-1^4).$$

The expression on the left $\sim \frac{n^6}{9}$. But the sum of fourth powers on the right $\sim \frac{n^5}{5}$, which is of lower order. Hence the sum on the right $\sim 2\pi_2$. We then have

$$\begin{aligned}\{1^2 + 2^2 + \ldots (n-1)^2\}^3 &\sim 2\pi_2\{1^2 + \ldots (n-1)^2\} \\ &\sim 6\pi_3 + \text{terms of type } 1^4 . 2^2.\end{aligned}$$

These terms will be less in sum than

$$2\{1^2 + 2^2 + \ldots (n-1)^2\}\{1^4 + 2^4 + \ldots (n-1)^4\}$$

which $\sim 2 . \frac{n^3}{3} . \frac{n^5}{5}$, of degree 8. But the expression on the left is of degree 9. Hence

$$\{1^2 + 2^2 + \ldots (n-1)^2\}^3 \sim 6\pi_3, \text{ and so on.}$$

We can now justify the assertion that the maximum power of n arises from terms like $(1^2 . 2^2 \ldots \alpha^2)$. In fact, by a similar line of reasoning to that just given it will appear that sums of terms like $(1^4 . 2^2 \ldots (\alpha - 1)^2)$ are of lower degree in n. This completes the demonstration.

16.23. In using the normal distribution to approximate to the S-distribution it is desirable to make a correction for continuity by subtracting unity (half the interval) from S in order to obtain the probability that a given value will be attained or exceeded. For instance, when $n = 9$ we have from (16.35)

$$\text{var } S = \frac{9 . 8 . 23}{18} = 92.$$

The normal deviate corresponding to $S = 20$ is then $\frac{19}{\sqrt{92}} = 1{\cdot}981$. The probability of a normal deviate as great as or greater than this is 0·0238. The value from Table 16.4 is 0·022. Had we made no correction for continuity we should have found a normal deviate of 2·085 with a probability of 0·0185.

Example 16.4

In **16.6** we found for a certain ranking of 10, $\tau = 0{\cdot}111$, $S = 5$. The Spearman coefficient for the same ranking, 0·139, has already been seen to be non-significant. What conclusion should we reach about τ on this point?

From Table 16.4 it is seen that the probability of a deviation greater than or equal to 5 is 0·364, and that of a deviation greater than or equal to 5 in absolute value is then 0·73 approximately. The corresponding value for ρ is 0·70. In either case the coefficient could well have arisen from an "independent" population and is not significant.

16.24. Different as ρ and τ are in conception and method of calculation, they are very closely related. Cogent reasons (but not a rigorous proof) have been given for belief in the validity of the equation (for the population in which all rankings occur equally frequently)—

$$\text{cov } (\rho, \tau) = -\frac{1}{18} n(n+1)^2(n-1)$$

from which the product–moment correlation between ρ and τ is

$$\frac{2(n+1)}{\sqrt{\{2n(2n+5)\}}} \qquad 1 - \frac{1}{4n} \tag{16.40}$$

(Kendall and others, 1938). For values of n occurring in practice the correlation between ρ and τ is thus very high. It also appears that the regression of ρ on τ is approximately linear over the material part of the range, that is, unless both are very close to unity. In such a case, recalling the values of the variances of the two coefficients, we shall have

$$\rho = \sqrt{\frac{1}{n-1}}.\tau\sqrt{\frac{18n(n-1)}{4(2n+5)}}$$

$$\doteq \frac{3\tau}{2},$$

so that τ will be about two-thirds of the value of ρ when n is large.

Grades

16.25. Up to this point we have considered the problem of rank correlation without reference to any variate system which might underlie the rankings. In certain classes of inquiry this is inevitable ; for example, we might shuffle a pack of cards and use the rank correlation between the orders before and after shuffling to measure the efficacy of the process of mixing. The early theory of rank correlation was, however, developed from rather a different view-point. The qualities considered were measurable, and always in theory (and often in practice) it was possible to find a product–moment coefficient of correlation. The use of Spearman's ρ was regarded as a substitute for such a coefficient, suitable either because the necessary measurements could not be carried out, whereas the ranking could, or because time was saved in working out rank correlations.

It is not immediately evident what meaning can be attached to ranking in a continuous population, for the members thereof are not denumerable.

The remark of **16.5** offers one way of overcoming the difficulty. The ranking of an individual as r can be regarded as a numerical statement to the effect that there are $(r - 1)$ members " above " that individual, that is to say $(r - 1)$ members who are given precedence. Quantities have already been considered in connection with continuous populations which express the same idea, namely, the quantiles. The pth decile, for example, is the variate-value such that p tenths of the total frequency lie below it. We will then define the *grade* of an individual as the proportion of the total frequency with a lower variate-value than that borne by that individual. If we have a discontinuous population N in number, the grade of an individual ranked according to the variate-values as r (from the lower to the higher values) will be $\frac{(r-1)}{N}$. If the population is continuous its members cannot be ranked ; but if we choose a sample of n members and rank them, an estimate of the grade of the rth member may be obtained by assuming that one-half of that member is to be assigned to each of the ranges into which its variate-value divides the variate-range, so that its grade is then taken to be

$$\frac{\{(r-1)+\frac{1}{2}\}}{n} = \frac{(r-\frac{1}{2})}{n}.$$

16.26. For a continuous bivariate population there will be no rank correlation, but there will, in general, be a grade correlation. Consider the bivariate normal population whose frequency function is

$$z = \frac{1}{2\pi(1-\rho'^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho'^2)}(x^2 - 2\rho'xy + y^2)\right\} \quad . \quad . \quad (16.41)$$

where, to avoid confusion with Spearman's ρ, we have denoted the product-moment coefficient by ρ'.

Let

$$\left.\begin{aligned}\xi &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} z\,dx\,dy = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{x} e^{-\frac{1}{2}x^2}\,dx\\ \eta &= \int_{-\infty}^{\infty}\int_{-\infty}^{y} z\,dy\,dx = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\,dy\end{aligned}\right\} \quad . \quad (16.42)$$

Then ξ and η are the grades and if x and y are independent so are ξ and η. ξ is a function of x and is distributed in the form

$$dF(\xi) = d\xi, \quad 0 \leqslant \xi \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad (16.43)$$

and similarly for η. Thus the mean and variance of both ξ and η are $\frac{1}{2}$ and $\frac{1}{12}$ respectively. For the Spearman coefficient between ξ and η we may then take

$$\rho = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi\eta\, z\,dx\,dy \quad . \quad . \quad . \quad . \quad . \quad (16.44)$$

remembering, however, that this is a generalisation of ρ to grades. From (16.44) we then have

$$\frac{d\rho}{d\rho'} = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi\eta\,\frac{\partial z}{\partial\rho'}\,dx\,dy.$$

Now $\qquad \log z = -\dfrac{1}{2(1-\rho'^2)}(x^2 - 2\rho'xy + y^2) - \frac{1}{2}\log(1-\rho'^2) - \text{constant.}$

Thus $\qquad \dfrac{1}{z}\dfrac{\partial z}{\partial\rho'} = \dfrac{\rho'}{(1-\rho'^2)^2}(x^2 - 2\rho'xy + y^2) + \dfrac{xy}{1-\rho'^2} + \dfrac{\rho'}{1-\rho'^2}$

$$\frac{1}{z}\,\frac{\partial^2 z}{\partial x\,\partial y}$$

and hence $\qquad \dfrac{d\rho}{d\rho'} = 12\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi\eta\,\frac{\partial^2 z}{dx\,dy}\,dx\,dy.$

By a partial integration with respect to x this is equal to

$$12\int_{-\infty}^{\infty} dy\left[\xi\eta\,\frac{\partial z}{\partial y}\right]_{-\infty}^{\infty} - 12\int_{-\infty}^{\infty} dy\int_{-\infty}^{\infty} dx\,\frac{\partial(\xi\eta)}{\partial x}\,\frac{\partial z}{\partial y}.$$

The first term vanishes and thus

$$\frac{d\rho}{d\rho'} = -12\int_{-\infty}^{\infty} dy\int_{-\infty}^{\infty} \eta\,\frac{\partial\xi}{\partial x}\,\frac{\partial z}{\partial y}\,dx.$$

By a partial integration with respect to y we find

$$\frac{d\rho}{d\rho'} = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial\eta}{\partial y}\,\frac{\partial\xi}{\partial x}\,z\,dx\,dy,$$

whence, from (16.42)

$$\frac{d\rho}{d\rho'} = \frac{12}{4\pi^2(1-\rho'^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{(2-\rho'^2)x^2 - 2\rho'xy + (2-\rho'^2)y^2}{2(1-\rho'^2)}\right\}dx\,dy = \frac{}{\pi(4-\rho'^2)^{\frac{1}{2}}}$$

Integrating we have, since ρ vanishes with ρ',

$$\rho = \frac{6}{\pi} \sin^{-1} \frac{\rho'}{2}$$

or

$$\rho' = 2 \sin \pi\rho \tag{16.45}$$

16.27. This formula is due to K. Pearson, but its value is problematical. It represents the relationship between the product–moment and the grade correlations when the variates are normal. It has, however, been used to transform a rank correlation obtained from a small sample of n values into a putative product–moment coefficient in that sample, or even worse, in the population from which the sample is derived, whether normal or not. The reader may care to list for himself the assumptions made in adopting such a procedure and to reflect on their justification. We shall not notice the process again, but we may note that in no case is ρ very different from $2 \sin \frac{\pi\rho}{6}$ in numerical value. If $\rho = 0{\cdot}6$, $2 \sin \frac{\pi\rho}{6} = 0{\cdot}618$, and this is about the greatest difference that can occur.

16.28. Equation (16.45) has also been advocated as an easy, though perhaps inaccurate, method of calculating a product–moment coefficient. The idea is that when a set of bivariate values is given they shall be replaced by ranks, the rank coefficient calculated, and the value of ρ' derived from (16.45). Apart from the theoretical objections, such a procedure involves no saving of labour if the number of values is greater than 30 or 40. Various formulae have been offered for the standard error of an estimate of the parent product–moment correlation based on (16.45). Some of those in current statistical text-books are incorrect, and it may be doubted whether the use of any one is justified. The reader may consult Eells (1929) for a list of these formulae.

The Case of m *Rankings*

16.29. We now consider the more general case in which there are m rankings of n instead of two. Our problem is to discuss the general agreement among the set of m.

It is natural in the first instance to consider the average ρ or τ in the $\binom{m}{2}$ possible pairs which can be chosen from the set of m. For example, if we have three rankings of six as follows : —

P	5	1	6	3	2
Q	2	1	5	6	4
R	4	6	3	2	5

(16.46)

the Spearman ρ's between PQ, QR and RP respectively are $\frac{11}{35}$, $-\frac{19}{35}$, $-\frac{19}{35}$, so that the average ρ, say ρ_{av}, is equal to $-\frac{9}{35} = -0{\cdot}26$. We shall consider a slightly different coefficient linearly related to ρ_{av}.

Suppose we sum the ranks in the columns of (16.46), obtaining the numbers

11 14 11 11.

These numbers must sum to 63 $\left(\text{and in general to } \frac{mn(n+1)}{2}\right)$ and reflect the degree of resemblance among the rankings. If the concordance were perfect the sums would be 3, 6, 9, 12, 15, 18, though not necessarily, of course, in that order, and in such a case would be as different as possible. On the other hand, when there is little or no resemblance, as in the example given, the sums are approximately equal. It is thus natural to take the variance of these sums as providing a measure of the ranking concordance.

Let S be the sum of the squares of deviations from the mean $\frac{m(n+1)}{2}$. If the concordance is perfect the sums are $m, 2m, \ldots nm$ and the sum S is $\frac{m^2(n^3-n)}{12}$. Write then

$$W = \frac{12S}{m^2(n^3-n)}. \tag{16.47}$$

Then W may vary from 0 to 1 and we shall call it the coefficient of concordance. In the above example it will be found that $S = 25{\cdot}5$, $W = 0{\cdot}16$.

16.30. W is connected with ρ_{av} by the relation

$$\rho_{av} = \frac{mW-1}{m-1}. \tag{16.48}$$

In fact, if the rankings, measured from the mean $\frac{1}{2}(n+1)$, are $x_{11}\, x_{12} \ldots x_{1n}, x_{21} \ldots x_{2n}, \ldots, x_{m1} \ldots x_{mn}$, the average ρ is

$$\frac{1}{m(m-1)}\cdot\frac{12}{n^3-n}\sum_{k,i=1}^{m}\sum_{j=1}^{n} x_{ij}\, x_{kj}, \qquad i \neq k\ . \tag{16.49}$$

$$\frac{12}{m(m-1)(n^3-n)}\left\{\sum_{j=1}^{n}\left(\sum_{i=1}^{m} x_{ij}\right)^2 - \sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij}^2\right\}$$

$$= \frac{12}{m(m-1)(n^3-n)}\left\{S - m.\frac{n^3-n}{12}\right\}$$

$$= \frac{mW-1}{m-1}.$$

ρ_{av} is the intra-class correlation for the m sets of ranks considered as variate-values. It cannot be less than $\frac{-1}{(m-1)}$.

16.31. To test whether an observed value of W is significant it is necessary to consider the distribution of W (or, more conveniently, of S) in the population obtained by permuting the n ranks in all possible ways in each of the m rankings. No generality is lost in supposing one ranking fixed and the others will then give rise to $(n!)^{m-1}$ values of S. We will ascertain the distributions for some low values of n and m and show how to approximate for larger values by the use of a continuous distribution.

For the case $m = 2$ the distribution of S is that of Σ in Table 16.1. The distributions have also been found for $n = 3$, $m = 2$ to 10; $n = 4$, $m = 2$ to 6; and $n = 5$, $m = 3$. Tables 16.5 to 16.8 give the probabilities based on these distributions in a form analogous to Tables 16.2 and 16.4.

TABLE 16.5

Concordance Coefficient W. Probability that a given Value of S will be Attained or Exceeded for $n = 3$ and Values of m from 2 to 10.

Values of m

S	2	3	4	5	6	7	8	9	10
0	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
2	0·833	0·944	0·931	0·954	0·956	0·964	0·967	0·971	0·974
6	0·500	0·528	0·653	0·691	0·740	0·768	0·794	0·814	0·830
8	0·167	0·361	0·431	0·522	0·570	0·620	0·654	0·685	0·710
14		0·194	0·273	0·367	0·430	0·486	0·531	0·569	0·601
18		0·028	0·125	0·182	0·252	0·305	0·355	0·398	0·436
24			0·069	0·124	0·184	0·237	0·285	0·328	0·368
26			0·042	0·093	0·142	0·192	0·236	0·278	0·316
32			0·0046	0·039	0·072	0·112	0·149	0·187	0·222
38				0·024	0·052	0·085	0·120	0·154	0·187
42				0·0085	0·029	0·051	0·079	0·107	0·135
50				$0{\cdot}0^{3}77$	0·012	0·027	0·047	0·069	0·092
54					0·0081	0·021	0·038	0·057	0·078
56					0·0055	0·016	0·030	0·048	0·066
62					0·0017	0·0084	0·018	0·031	0·046
72					$0{\cdot}0^{3}13$	0·0036	0·0099	0·019	0·030
74						0·0027	0·0080	0·016	0·026
78						0·0012	0·0048	0·010	0·018
86						$0{\cdot}0^{3}32$	0·0024	0·0060	0·012
96						$0{\cdot}0^{3}32$	0·0011	0·0035	0·0075
98						$0{\cdot}0^{4}21$	$0{\cdot}0^{3}86$	0·0029	0·0063
104							$0{\cdot}0^{3}26$	0·0013	0·0034
114							$0{\cdot}0^{4}61$	$0{\cdot}0^{3}66$	0·0020
122							$0{\cdot}0^{4}61$	$0{\cdot}0^{3}35$	0·0013
126							$0{\cdot}0^{4}61$	$0{\cdot}0^{3}20$	$0{\cdot}0^{3}83$
128							$0{\cdot}0^{5}36$	$0{\cdot}0^{4}97$	$0{\cdot}0^{3}51$
134								$0{\cdot}0^{4}54$	$0{\cdot}0^{3}37$
146								$0{\cdot}0^{4}11$	$0{\cdot}0^{3}18$
150								$0{\cdot}0^{4}11$	$0{\cdot}0^{3}11$
152								$0{\cdot}0^{4}11$	$0{\cdot}0^{4}85$
158								$0{\cdot}0^{4}11$	$0{\cdot}0^{4}44$
162								$0{\cdot}0^{6}60$	$0{\cdot}0^{4}20$
168									$0{\cdot}0^{4}11$
182									$0{\cdot}0^{5}21$
200									$0{\cdot}0^{7}99$

TABLE 16.6

Concordance Coefficient W. Probability that a given Value of S will be Attained or Exceeded for $n = 4$ and $m = 3$ and 5.

S	$m = 3$	$m = 5$	S	$m = 5$
1	1·000	1·000	61	0·055
3	0·958	0·975	65	0·044
5	0·910	0·944	67	0·034
9	0·727	0·857	69	0·031
11	0·608	0·771	73	0·023
13	0·524	0·709	75	0·020
17	0·446	0·652	77	0·017
19	0·342	0·561	81	0·012
21	0·300	0·521	83	0·0087
25	0·207	0·445	85	0·0067
27	0·175	0·408	89	0·0055
29	0·148	0·372	91	0·0031
33	0·075	0·298	93	0·0023
35	0·054	0·260	97	0·0018
37	0·033	0·226	99	0·0016
41	0·017	0·210	101	0·0014
43	0·0017	0·162	105	$0{\cdot}0^{3}64$
45	0·0017	0·141	107	$0{\cdot}0^{3}33$
49		0·123	109	$0{\cdot}0^{3}21$
51		0·107	113	$0{\cdot}0^{3}14$
53		0·093	117	$0{\cdot}0^{4}48$
57		0·075	125	$0{\cdot}0^{5}30$
59		0·067		

TABLE 16.7

Concordance Coefficient W. Probability that a given Value of S will be Attained or Exceeded for n = 4 and m = 2, 4 and 6.

S	$m = 2$	$m = 4$	$m = 6$	S	$m = 6$
0	1·000	1·000	1·000	82	0·035
2	0·958	0·992	0·996	84	0·032
4	0·833	0·928	0·957	86	0·029
6	0·792	0·900	0·940	88	0·023
8	0·625	0·800	0·874	90	0·022
10	0·542	0·754	0·844	94	0·017
12	0·458	0·677	0·789	96	0·014
14	0·375	0·649	0·772	98	0·013
16	0·208	0·524	0·679	100	0·010
18	0·167	0·508	0·668	102	0·0096
20	0·042	0·432	0·609	104	0·0085
22		0·389	0·574	106	0·0073
24		0·355	0·541	108	0·0061
26		0·324	0·512	110	0·0057
30		0·242	0·431	114	0·0040
32		0·200	0·386	116	0·0033
34		0·190	0·375	118	0·0028
36		0·158	0·338	120	0·0023
38		0·141	0·317	122	0·0020
40		0·105	0·270	126	0·0015
42		0·094	0·256	128	$0{\cdot}0^{3}90$
44		0·077	0·230	130	$0{\cdot}0^{3}87$
46		0·068	0·218	132	$0{\cdot}0^{3}73$
48		0·054	0·197	134	$0{\cdot}0^{3}65$
50		0·052	0·194	136	$0{\cdot}0^{3}40$
52		0·036	0·163	138	$0{\cdot}0^{3}36$
54		0·033	0·155	140	$0{\cdot}0^{3}28$
56		0·019	0·127	144	$0{\cdot}0^{3}24$
58		0·014	0·114	146	$0{\cdot}0^{3}22$
62		0·012	0·108	148	$0{\cdot}0^{3}12$
64		0·0069	0·089	150	$0{\cdot}0^{4}95$
66		0·0062	0·088	152	$0{\cdot}0^{4}62$
68		0·0027	0·073	154	$0{\cdot}0^{4}46$
70		0·0027	0·066	158	$0{\cdot}0^{4}24$
72		0·0016	0·060	160	$0{\cdot}0^{4}16$
74		$0{\cdot}0^{3}94$	0·056	162	$0{\cdot}0^{4}12$
76		$0{\cdot}0^{3}94$	0·043	164	$0{\cdot}0^{5}80$
78		$0{\cdot}0^{3}94$	0·041	170	$0{\cdot}0^{5}24$
80		$0{\cdot}0^{4}72$	0·037	180	$0{\cdot}0^{6}13$

TABLE 16.8

Concordance Coefficient W. Probability that a given Value of S will be Attained or Exceeded for $n = 5$ and $m = 3$.

S	$m = 3$	S	$m = 3$
0	1·000	44	0·236
2	1·000	46	0·213
4	0·988	48	0·172
6	0·972	50	0·163
8	0·941	52	0·127
10	0·914	54	0·117
12	0·845	56	0·096
14	0·831	58	0·080
16	0·768	60	0·063
18	0·720	62	0·056
20	0·682	64	0·045
22	0·649	66	0·038
24	0·595	68	0·028
26	0·559	70	0·026
28	0·493	72	0·017
30	0·475	74	0·015
32	0·432	76	0·0078
34	0·406	78	0·0053
36	0·347	80	0·0040
38	0·326	82	0·0028
40	0·291	86	$0·0^{3}90$
42	0·253	90	$0·0^{4}69$

16.32. These distributions may be obtained by two methods. The first consists of building up the distribution for $(m + 1)$ and n from that for m and n. For example, with $m = 2$ and $n = 3$ we have the following values of the sums of ranks, measured about their mean :—

Type			Frequency
-2	0	2	1
-2	1	1	2
-1	0	1	2
0	0	0	1

Here $-2, 1, 1$, and $2, -1, -1$ are taken to be identical types, for they give the same value of S and will also give similar types when we proceed to $m = 3$ as follows.

In the case $m = 3$, each of the above type will appear added to the six permutations of $-1, 0, 1$; e.g. the type $-2, 0, 2$ will give one each of $-3, 0, 3$; $-3, 1, 2$; $-2, -1, 3$; $-2, 1, 1$; $-1, -1, 2$; and $-1, 0, 1$. These types are then counted for each of the four basic types of $m = 2$ and we get :—

Type			Frequ
-3	0	3	1
-3	1	2	6
-2	0	2	6
-2	1	1	6
-1	0	1	15
0	0	0	2

The case $m = 4$ is treated by considering the numbers of types obtained by adding the six permutations of -1, 0, 1 to the types for $m = 3$; and so on.

This method is quite convenient for $n = 2$ and $n = 3$. For $n = 4$ it becomes difficult owing to the labour of considering 24 permutations at each stage and to the increase in the number of types. For $n = 5$ there are 120 permutations and the labour becomes excessive.

The second method is a generalisation of the E-function of **16.12.** For m rankings, the distribution of S is given by the expansion of an m-dimensional E-function. For example, with $m = 3$ there would be a three-dimensional E-function the bottom plane of which would be

$$\begin{array}{cccc}
a^{\left\{3-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & & a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} \\
a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{5-\frac{3(n+1)}{2}\right\}^2} & . & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\
\dots & \dots & \dots & \dots \\
a^{\left\{n+2-\frac{3(n+1)}{2}\right\}^2} & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{2n+2-\right.}
\end{array}$$

The plane above this would be

$$\begin{array}{ccc}
a^{\left\{4-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} \\
\dots & \dots & \dots \\
a^{\left\{n+3-\frac{3(n+1)}{2}\right\}^2} & \dots & a^{\left\{2n+3-\frac{3(n+1)}{2}\right\}^2}
\end{array}$$

and so on.

The E-function is difficult to handle in more than three dimensions, but for the two- and three-dimensional case it is manageable and was used to obtain the distribution of S for $n = 5$ and $m = 3$.

16.33. We now proceed to find the first four moments of the distribution of S. The method is similar to that used for ρ but is somewhat more complicated.

Writing x_{ij} for the deviation from the mean $\frac{(n+1)}{2}$ of the jth member of the ith ranking, we have, as in **16.30,**

$$W = \frac{1}{m} + \frac{m-1}{m}\rho_a$$

$$= \frac{1}{m} + \frac{1}{m^2}\cdot\frac{12}{n^3-n}\cdot\sum_{i,k=1}^{m}\sum_{j=1}^{n} x_{ij}\,x_{kj}, \qquad i \neq k \tag{16.50}$$

Write
$$R_{ik} = \sum_{j=1}^{n} x_{ij}\,x_{kj} \tag{16.51}$$

where i, k can have all values from 1 to m and thus any term $R_{\alpha\beta}$ appears again as $R_{\beta\alpha}$ in the sum ΣR_{ik}. Then the moments of W are derivable from those of the R's, which

in turn are derivable from those of Spearman's ρ. In fact, writing $N = \frac{(n^3 - n)}{12}$ we have from (16.18) and (16.19)

$$\begin{aligned} E(R_{ik}) &= 0,\ E(R_{ik}^3) = 0 \\ E(R_{ik}^2) &= N^2 . \frac{1}{n-1} \\ E(R_{ik}^4) &= N^4 \left\{ \frac{3(25n^3 - 38n^2 - 35n + 72)}{25n(n+1)(n-1)^3} \right\} \end{aligned} \qquad (16.52)$$

We next require the moments of

$$\rho_{av} = \frac{1}{m(m-1)N} \sum_{i,l} R_{ik}$$

but complications arise because in some cases the R's are correlated among themselves. Any two R's are independent, i.e.

$$E(R_{ik}\ R_{lm}) = 0, \qquad (16.53)$$

unless of course $i = l$, $k = m$. This may be seen by reference to (16.51), the x's being independent. Similarly

$$E(R_{ik}\ R_{lm}\ R_{np}) = 0,$$

except when we have a set with "circular" suffixes such as

$$E(R_{ik}\ R_{kl}\ R_{li}), \qquad (16.54)$$

for in this case the x's cease to be independent. Similarly any four R's are independent unless they form a set such as

$$R_{ik}\ R_{kl}\ R_{lm}\ R_{mi}. \qquad (16.55)$$

We have

$$\begin{aligned} E(R_{ik}\ R_{kl}\ R_{li}) &= E\left(\sum_{\alpha=1}^{n} x_{i\alpha}\, x_{k\alpha} \sum_{\beta=1}^{n} x_{k\beta}\, x_{l\beta} \sum_{\gamma=1}^{n} x_{l\gamma}\, x_{i\gamma} \right) \\ &= E\left[\{\Sigma x_{k\alpha}^2\, x_{i\alpha}\, x_{l\alpha} + \Sigma\, x_{k\alpha}\, x_{k\beta}(x_{i\alpha}\, x_{l\beta} + x_{i\beta}\, x_{l\alpha})\} \{\Sigma\, x_{l\gamma}\, x_{i\gamma}\} \right] \\ &= E\left[\{E(x_{k\alpha}^2)\, \Sigma\, x_{i\alpha}\, x_{l\alpha}\} + E(x_{k\alpha}\, x_{k\beta}) \{\Sigma x_{i\alpha}\, x_{l\alpha} - \Sigma x_{i\alpha}\, \Sigma x_{l\alpha}\} \right] \times [\Sigma\, x_{l\gamma}\, x_{i\gamma}] \\ &= \{E(x_{k\alpha}^2) - E(x_{k\alpha}\, x_{k\beta})\}\, E(\Sigma x_{i\alpha}\, x_{l\alpha})^2 \\ &= N\left\{ \frac{1}{n} + \frac{1}{n(n-1)} \right\} N^2 . \frac{1}{n-1} \\ &= \frac{N^3}{(n-1)^2} \end{aligned} \qquad (16.56)$$

We then have

$$\begin{aligned} E(\rho_{av}) &= \frac{1}{m(m-1)N} E\{\Sigma(R_{ik})\} \\ &= \frac{1}{m(m-1)N} \Sigma\{E(R_{ik})\} \\ &= 0 \end{aligned} \qquad (16.57)$$

$$E(\rho_{av}^2) = \frac{1}{m^2(m-1)^2N^2} E(\Sigma R_{ik})^2$$

$$= \frac{1}{m^2(m-1)^2N^2} E(\Sigma R_{ik}^2 + \Sigma R_{ki} R_{ik} + \Sigma R_{ik} R_{il})$$

$$= \frac{1}{m^2(m-1)^2N^2} E(2\Sigma R_{ik}^2)$$

$$= \frac{2}{m^2(m-1)^2N^2} . m(m-1) \frac{N^2}{n-1}$$

$$= \frac{2}{m(m-1)} . \frac{1}{n-1} \qquad (16.58)$$

$$E(\rho_{av}^3) = \frac{1}{m^3(m-1)^3N^3} E(\Sigma R_{ik})^3$$

$$= \frac{1}{m^3(m-1)^3N^3} E\, \Sigma(R_{ik} R_{kl} R_{li}),$$

all other terms vanishing,

$$= \frac{8m(m-1)(m-2)}{m^3(m-1)^3N^3} E(R_{ik} R_{kl} R_{li})$$

$$= \frac{8(m-2)}{m^2(m-1)^2} . \frac{1}{(n-1)^2} . \qquad (16.59)$$

From these results we have, for the first three moments of W,

$$\mu_1' \text{ (about 0)} = \frac{1}{m} \quad . \quad . \quad . \qquad (16.60)$$

$$\mu_2 = \frac{2(m-1)}{m^3(n-1)} \quad . \qquad (16.61)$$

$$\mu_3 = \frac{8(m-1)(m-2)}{m^5(n-1)^2} \quad . \quad . \quad . \qquad (16.62)$$

In a similar way—we omit the lengthy algebra—it may be shown that

$$\mu_4 = \frac{24(m-1)}{m^7(n-1)^2} \left\{ \frac{25n^3 - 38n^2 - 35n + 72}{25(n^3 - n)} + 2(n-1)(m-2) + \tfrac{1}{2}(n+3)(m-2)(m-3) \right\} . \qquad (16.63)$$

16.34. The distribution of W is evidently asymmetrical since $\mu_3 \neq 0$ unless $m = 2$. Consider then the possibility of approximating to the distribution by the Type I form

$$dF = \frac{1}{B(p, q)} W^{p-1}(1-W)^{q-1}\, dW, \qquad 0 \leqslant W \leqslant 1 \qquad (16.64)$$

The first two moments of this are

$$\mu_1' \text{ (about 0)} = \frac{p}{p+q}$$

$$\mu_2 = \frac{pq}{(p+q)^2(p+q+1)} \qquad (16.65)$$

Identifying the values of (16.60), (16.61) and (16.65), we find

$$\frac{p}{p+q} = \frac{1}{m}$$

$$\frac{pq}{(p+q)^2(p+q+1)} = \frac{2(m-1)}{m^3(n-1)}$$

giving

$$\left.\begin{aligned} p &= \tfrac{1}{2}(n-1) - \frac{1}{m} \\ q &= (m-1)\left\{\tfrac{1}{2}(n-1) - \frac{1}{m}\right\} \end{aligned}\right\} \qquad (16.66)$$

It will be found that the third moment about the mean of the Type I form is

$$\frac{8(m-1)(m-2)}{m^4(n-1)(mn+m-2)} = \frac{8(m-1)(m-2)}{m^5(n-1)}\left\{1 - \frac{}{m(n-1)+2}\right\}$$

so that the third moments of the W-distribution and the Type I distribution are approximately equal if m and n are not small. Similarly the fourth moments will be found to differ by a small quantity. We may therefore use the Type I distribution to approximate to that of W. It appears likely that as $n, m \to \infty$ the distribution of W tends to the Type I form, but this has not been rigorously demonstrated.

16.35. The significance of W can then be tested in the Type I distribution, namely, by the use of incomplete B-functions. More conveniently, we may transform (16.64) to the form

$$dF = k \frac{e^{\nu_1 z}}{(\nu_1 e^{2z} + \nu_2)^{\frac{\nu_1+\nu_2}{2}}} dz \qquad (16.67)$$

by the transformation

$$z = \tfrac{1}{2} \log \frac{(m-1)W}{1-W}$$

$$\nu_1 = (n-1) - \frac{2}{m}$$

$$\nu_2 = (m-1)\left\{(n-1) - \frac{}{m}\right.$$

and test in the z-distribution which has been tabulated.

In making this test it is desirable, for low values of m and n, to make the usual correction for continuity by subtracting unity from S (equation (16.47)) and increasing the divisor $\frac{m^2(n^3-n)}{12}$ by 2. Let us examine the approximation of the test in some cases wherein the exact values are known from Tables 16.5 to 16.8.

For $n = 3$, $m = 9$, the 1 per cent. level is given approximately by $S = 78$ (Table 16.5). For such a value, with continuity corrections,

$$W = \frac{(78-1)}{\frac{81.24}{12} + 2} = 0{\cdot}4695$$

$$z = 0{\cdot}979$$

$$\nu_1 = \frac{16}{9}, \quad \nu_2 = 128$$

By linear interpolation of reciprocals in Appendix Table 5 we should require, for complete agreement, a value of z equal to 0·954.

For $n = 4$, $m = 6$, the 1 per cent. point is approximately $S = 100$. $\nu_1 = 8/3$, $\nu_2 = 40/3$. We have $W = 0{\cdot}5556$, $z = 0{\cdot}916$. From Table 16.5 we should require a value of 0·893.

For $n = 5$, $m = 3$, there is no very convenient value of S close to the 1 per cent. point. For $P = 0{\cdot}015$, $S = 74$, and for $P = 0{\cdot}0078$, $S = 76$.

For $S = 74$ (with continuity corrections) $z = 1{\cdot}020$
$S = 76$ (,, ,, ,,) $z = 1{\cdot}089$.

By interpolation from the tables $z = 1{\cdot}075$. The use of the z-test would lead to the correct conclusion that a value of S equal to 74 falls below, and that of 76 above, the 1 per cent. point.

For values of m and n not included in Tables 16.5 to 16.8 it thus appears that the z-test with continuity corrections will give sufficiently accurate results, if n is greater than 3, at the 1 per cent. points. It may be presumed that the results at the 5 per cent. points are equally good and probably better. But for finer values of significance, such as 0·1 per cent., it is doubtful whether the test is sound. The tails of the distribution of S for moderate values of m and n are very irregular.

16.36. A somewhat more approximate test of W has been given by Friedman (1937), who defined a statistic

$$\chi_r^2 = m(n-1)W \qquad . \quad . \quad . \quad . \quad . \quad (16.68)$$

and showed that the distribution of χ_r^2 tends to that of the Type III χ^2 as n tends to infinity, with $(n-1)$ degrees of freedom. This test appears reasonably satisfactory for moderate m and n, though not so accurate as ours. Friedman has also provided (1940) some significance levels of χ_r^2 calculated on the basis of the z-test.

Example 16.5

In some experiments in random series a pack of ordinary playing cards was shuffled and the order of the 13 cards of each suit from the top of the pack was noted. The pack was then reshuffled and again the orders noted. This was done 28 times. The question to discuss was whether the shuffling was good, in the sense that the cards were thoroughly mixed at each shuffle.

Here, for each suit, say diamonds, we have 28 rankings of 13. The sums of ranks were 183, 137, 171, 207, 188, 160, 225, 174, 216, 192, 236, 239, 220. The mean is 196, and $S = 11{,}522$; W (without continuity corrections, which are not worth making for these values of m and n) $= 0{\cdot}08075$, $z = 0{\cdot}432$. This falls just beyond the 1 per cent. point.

Similarly for the clubs W was found to be 0·0535; for the hearts, 0·0245; and for spades, 0·0342. None of these values is significant, and we conclude that the randomisation introduced by the shuffling was good, at all events, so far as this test was concerned. It may be added that the shuffling was done with much more care than would be taken in an ordinary game of cards.

Example 16.6

In psychological work there has sometimes been a confusion between the determination of a measure of agreement between subjects and that of an objective order based on experimental rankings. It may therefore be as well to point out that in its psycho-

logical applications the test of W is one of concordance between judgments. There may be quite a high measure of agreement about something which is incorrect.

A number of students were given 12 photographs of persons unknown to them, and asked to rank them in what they judged from the photographs to be their intelligence. For 16 students the sums of ranks were

$$112, 94, 101, 84, 97, 75, 104, 84, 102, 146, 125, 124.$$

The mean is 104. $S = 4472$, $W = 0{\cdot}1222$. $z = 0{\cdot}368$, and is barely significant, being between the 1 per cent. and the 5 per cent. points.

For 111 students the sums were

$$818, 670, 908, 410, 706, 526, 780, 485, 596, 1044, 959, 756$$

$$W = 0{\cdot}2378, \ z = 1{\cdot}768.$$

This is highly significant and it is to be inferred that community of judgment exists between students or groups of students. But there was little relationship between the judgments and the intelligence of the photographed subjects as given by the Binet Intelligence Quotient.

Estimation of a True Ranking

16.37. Suppose we have m sets of n rankings which show a significant concordance. Assuming that the relations between the rankings reflect the true ranking of the objects, how are we to estimate that ranking ? or again, assuming merely a significant concordance between observers, what is the ranking " nearest " to their rankings ?

An intuitive approach to this problem would probably lead us to this solution : the object whose true rank is 1 is that for which the sum of ranks is least ; that whose rank is 2 is the one for which the sum of ranks is least but one ; and so on. For example, if there are three rankings of five objects totalling 9, 7, 4, 10, 15, we should take the third as rank 1, the second as 2, the first as 3, the fourth as 4 and the fifth as 5.

This solution can be given a firmer theoretical basis. It is the " best " in a least-squares sense. In fact, suppose the true ranking is $X_1, X_2 \ldots X_n$, where as usual the X's are a permutation of the first n integers. Suppose the sums of ranks are $S_1, S_2 \ldots S_n$.

Consider the sum

$$U = \Sigma(S_i - mX_i)^2. \qquad (16.69)$$

If all the rankings were correct, each S_i would be mX_i, so that this quantity represents in a sense the divergence from complete agreement. Our " best " estimate of the X's will be given by minimising U. Now

$$U = \Sigma(S_i^2) + m^2\,\Sigma(X_i^2) - 2m\,\Sigma\,S_i\,X_i$$

and since the first two terms are constant we have to maximise $\Sigma(S_i\,X_i)$, the S's being given and the X's the numbers 1 to n. Evidently this will be done by multiplying the biggest S by n, the next biggest by $(n - 1)$, and so on. The result follows.

There is, of course, an indeterminacy in this method if any two of the S's are equal.

Paired Comparisons

16.38. When the objects which are being ranked are known to be measurable according to the quality concerned, no question as to the legitimacy of ranking arises. But cases occur in which it is by no means clear that ranking is legitimate, as for instance in the arranging of human beings according to intelligence or of pieces of music by human beings according to preference. To require an observer to carry out a ranking in such a case

may be equivalent to asking him to arrange English towns in order of geographical position (which is two-dimensional), or a number of fruits according to taste (which is probably four-dimensional). The observer may attempt to comply in the full belief that he is doing something within his powers, but if the quality under consideration is not measurable on a linear scale the resulting ranking may fail to give either a real picture of his preferences or of the variation of the quality in the individuals. For example, in judgments of intelligence, it is not impossible that the observer should judge A more intelligent than B, B than C, and C than A, if the individuals are presented for his consideration one pair at a time. The likelihood of this happening is obviously increased when we are dealing with tastes in music, eatables or film stars ; and in practice the event is not uncommon. Such "inconsistent" preferences can never appear in ranking, for if A is preferred to B and B to C, then A must automatically be shown as preferred to C.

16.39. We therefore consider a more general method of investigating preferences. With n objects, we shall suppose that each of the $\binom{n}{2}$ possible pairs is presented to an observer and his preference of one member of the pair noted. If the object A is preferred to B we write $A \rightarrow B$ or $B \leftarrow A$. The $\binom{n}{2}$ preferences of a single observer may be represented in tabular form as shown in Table 16.9.

In this table, which is shown for the six objects A to F, an entry of unity in column Y and row X means $X \rightarrow Y$, and is thus accompanied by a complementary zero in row Y and column X. The diagonals are blocked out. For example, in the table, $A \rightarrow B$, $A \rightarrow C$, $D \rightarrow A$, etc.

TABLE 16.9

Tabular Representation of Paired Comparison Schema.

	A	B	C	D	E	F
A	—	1	1	0	1	1
B	0	—	0	1	1	0
C	0	1	—	1	1	1
D	1	0	0	—	0	0
E	0	0	0	1	—	1
F	0	1	0	1	0	—

The arrangement of the objects A to F in the row and column headings is quite arbitrary. There are $(n!)^2$ ways of representing the same configuration of preferences in such a table

according to the permutations of objects in row and column; but in practice it is generally desirable to have the order in row and column the same, and even among the $n!$ possible arrangements so given there are often practical considerations which determine one order as more convenient than others.

16.40. Paired comparisons may also be represented geometrically by a method which can be illustrated for the case of the six objects as follows :—

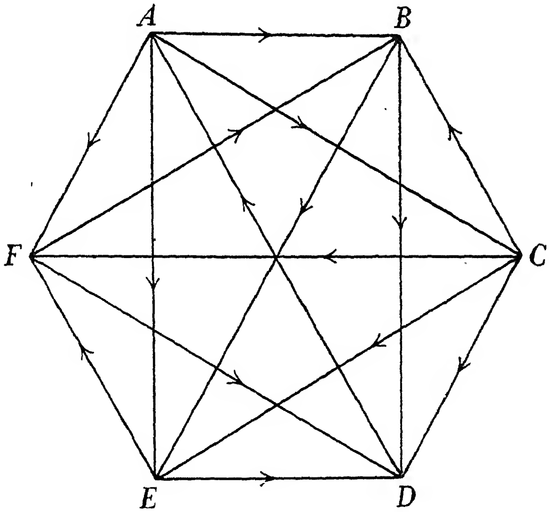


FIG. 16.2.—Geometrical Representation of the Scheme of Preferences of Table 16.9.

We represent the six objects A to F by the six vertices of a regular hexagon and join the vertices in all possible ways by straight lines. If $A \to B$ we draw an arrow on the line AB pointing from A to B. The arrows shown on Fig. 16.2 correspond to the preferences shown in Table 16.9.

16.41. If an observer makes preferences of type $A \to B \to C \to A$ we say that the triad ABC is inconsistent. In the geometrical representation an inconsistent triad is shown by a triangle in which all the arrows go round in the same direction. We may thus speak of a "circular" triad of preferences. In Fig. 16.2 the triads ACD, BEF and three others are circular.

It is also possible to have inconsistent triads of greater extent; but any such circuit must contain at least two circular triads. Suppose, for instance, that $ABCD$ is circular, e.g. that $A \to B \to C \to D \to A$. Then either $A \to C$ or $C \to A$. In the first case ACD is circular, in the second ABC. Similarly either ABD or BCD is circular. Thus the circular tetrad must contain just two circular triads. On the other hand it is possible for a tetrad to contain circular triads without being itself circular.

Similarly, if $ABCDE$ is circular either ABC or $ACDE$ is circular and either BCD or $BDEA$ is circular. If the two tetrads are circular there must be at least three circular triads (not necessarily four, because ADE may be common to both). It is easy to see by an actual example based on this configuration that there need not be more than three circular triads; and it is clear that there must be at least three. For if the tetrads are

not circular then ABC and BCD must be so and then either CDE is circular or $ABCE$ is so, adding at least one more.

Generally, it appears that a circular n-ad must contain at least $(n - 2)$ circular triads; but it may contain more, and the fact that an n-ad contains $(n - 2)$ circular triads does not mean that it is itself circular. In discussing inconsistences, therefore, it seems best to confine attention to circular triads, which, so to speak, constitute the inconsistent elements of the configuration, and to ignore the more ambiguous criteria associated with circular polyads of greater extent.

16.42. We now prove the following theorems :—

(1) The maximum possible number of circular triads is $\frac{n^3 - n}{24}$ if n is odd and $\frac{n^3 - 4n}{24}$ if n is even; and the minimum number is zero.

(2) These limits can always be attained by some configuration of preferences.

Consider a polygon of the type shown in Fig. 16.2 with n vertices. There will be $(n - 1)$ lines emanating from each vertex. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the number of lines at the respective vertices on which the arrows *leave* the vertex.

Then
$$\sum_{r=1}^{n} (\alpha_r) = \binom{n}{2}$$
and the mean value of α_r is $\frac{n-1}{2}$.

Define
$$T = \sum_{r=1}^{n} \left(\alpha_r - \frac{n-1}{2}\right)^2$$
$$= \Sigma(\alpha_r^2) - n(n-1)^2 \quad . \quad (16.70)$$

We now show that if the direction of a preference is altered and the effect is to increase the number of circular triads by d, T is reduced by $2d$; and conversely. Consider the preference $A \rightarrow B$. The only triads affected by altering this to $B \rightarrow A$ are those containing the line AB. Suppose there are α preferences of type $A \rightarrow X$ (including $A \rightarrow B$) and β preferences of type $B \rightarrow X$. Then four possible types of triad arise:

$A \dashrightarrow X \leftarrow B$, say p in number
$A \dashleftarrow X \rightarrow B$,
$A \dashrightarrow X \rightarrow B$, which must number $\alpha - p - 1$
$A \dashleftarrow X \leftarrow B$, ,, ,, ,, $\beta - p$.

When the preference $A \rightarrow B$ is reversed the first two remain non-circular. The third becomes circular, the fourth ceases to be so. The reduction in the value of T is

$$\begin{aligned} \alpha^2 - (\alpha - 1)^2 + \beta^2 - (\beta + 1)^2 &= 2(\alpha - \beta - 1) \\ &= 2d, \text{ say.} \end{aligned}$$

The increase in the number of circular triads is

$$\begin{aligned} (\alpha - p - 1) - (\beta - p) &= \alpha - \beta - 1 \\ &= d. \end{aligned}$$

More generally, if as the result of reversing any number of preferences T is decreased by $2\,d$, then d must be an integer and the number of circular triads must be increased by d. This clearly follows from the previous results, for the reversal of preferences can take place one at a time and the effect on T and the number of circular triads is cumulative.

We now investigate the maximum and minimum values of T. It is clear from the definition that T is greatest when the α's are the natural numbers $1, 2, \ldots, n$; and this is a possible case because it corresponds to ordinary ranking. Hence max. $(T) = \frac{n^3 - n}{12}$

For the minimum value, consider the polygon $A_1, A_2, \ldots, A_n$. Set up the preferences $A_1 \to A_2 \to \ldots A_n \to A_1$. Clearly at any vertex this results in one arrow entering and one leaving the vertex, i.e. the contribution to α is unity at each vertex. Next set up the preferences $A_1 \to A_3 \to A_5 \to \ldots$. This circuit may either visit each vertex once, or not. In the latter case we proceed to an unvisited vertex and set up the preferences $A_r \to A_{r+2} \to A_{r+4} \to \ldots$, and so on. Again there will be a unit contribution to all the α's.

We then set up the preferences $A_1 \to A_4 \to A_7 \to$, etc., and so on; and in this way we shall ultimately complete the preference scheme.

If n is odd all the preferences described will consist of circular tours of the polygon, and thus the value of α for each vertex will be $\frac{n-1}{2}$. If n is even, the last preference $A_1 \to A_{\frac{1}{2}n+1}$ will not be a tour but will consist of the single line joining one vertex with the symmetrically opposite vertex. Thus here will be $\frac{n}{2}$ vertices for which $\alpha = \frac{n}{2}$ and $\frac{n}{2}$ vertices for which $\alpha = \frac{n-2}{2}$. In this case $T = \frac{n}{4}$.

Now it is clear from the definition of T that it cannot be less than zero, or if n is even, be less than $\frac{n}{4}$. The configuration just given shows that these minima are, in fact, attainable.

Thus T can vary from a maximum of $\frac{n^3 - n}{12}$ to a minimum of zero or $\frac{n}{4}$. Hence the maximum number of circular triads, being half the variation from maximum to minimum of T (the maximum of T corresponding to the ranking case in which there are no inconsistences), is $\frac{n^3 - 4n}{24}$ if n is even and $\frac{n^3 - n}{24}$ if n is odd.

This establishes the two results enunciated at the beginning of this section.

Coefficient of Consistence in Paired Comparisons

16.43. If d is the number of circular triads in an observed configuration of preferences we define

$$\left.\begin{aligned} \zeta &= 1 - \frac{24d}{n^3 - n}, \quad n \text{ odd} \\ &= 1 - \frac{24d}{n^3 - 4n}, \quad n \text{ even} \end{aligned}\right\} \qquad . \ (16.71)$$

and call ζ the coefficient of consistence. If and only if it is unity, there are no inconsistences in the configuration, which may therefore be represented by a ranking. As ζ decreases to zero the inconsistence, as measured by the number of circular triads, increases.

For example, in the configuration of Fig. 16.2 there are five circular triads, ABD, ACD, AFD, AED and BEF. The maximum possible number is 8. Thus $\zeta = 0{\cdot}375$.

ζ can also be interpreted in the light of Table 16.9. Suppose, in that table, we sum the rows. (The column sums are determined by the row sums and add no fresh information.) The sum of any row will be the α-number for that vertex in the polygon which corresponds to the object defining the row. T will then be the value of the sum of squares of deviations of row totals from the mean value $\frac{n-1}{2}$, that is to say, will be the variance of the row sums multiplied by n. ζ is thus a linear function of this variance; but it cannot be tested in the χ^2-distribution as if Table 16.9 were a contingency table, for the border cells are not independent or linearly dependent.

16.44. If an individual observer produces a configuration of preferences which show inconsistence there are usually several explanations; he may be an incompetent judge, the objects may be so alike that consistent differentiation is not possible, or his attention may wander during the course of the experiment. We discuss these questions later. They are mentioned here to explain the motive for the next stage of the mathematics. With what probability can a value of ζ arise by chance if the observer allots his preferences at random with respect to the quality under consideration?

With n objects there are $2^{\binom{n}{2}}$ possible configurations of preferences. We proceed to investigate the distribution of d in this population of $2^{\binom{n}{2}}$ different members. The method consists of proceeding from the distribution for n to that for $(n + 1)$.

For $n = 3$ there are eight configurations, of which two give one circular triad and six no circular triads. Consider the effect of adding a new vertex D to the vertices ABC. Four cases arise:

(1) $D \rightarrow$ all A, B, C.
(2) $D \rightarrow$ two of A, B, C.
(3) $D \rightarrow$ one of A, B, C.
(4) $D \rightarrow$ none of A, B, C.

The last two are symmetrical with the first two and need not be separately considered.

Situation (1) arises in one way and clearly does not add any new circular triads other than those already existing in the configuration ABC. It therefore contributes six values $d = 0$ and two values $d = 1$. So does situation (4).

Situation (2) arises in three ways, according as $D \leftarrow A$, B, or C. The configurations so reached are similar and we may take any one, say $D \leftarrow C$, as the single preference. If $A \leftarrow C$ then DAC is not circular and if $B \leftarrow C$ then DBC is not circular. On the other hand $A \rightarrow C$ and $B \rightarrow C$ will each produce a circular triad. We then have the cases

	No. of Circular Triads added.
$A \leftarrow C \rightarrow B$	0
$A \rightarrow C \rightarrow B$	1
$A \leftarrow C \leftarrow B$	1
$A \rightarrow C \leftarrow B$	2

We now consider AB. In the first two cases just enumerated the direction of AB does not matter and no circular triads are added. With the third $A \rightarrow B$ gives no circular triad but $A \leftarrow B$ adds one. With the fourth $A \rightarrow B$ adds one and $A \leftarrow B$ adds none.

Thus the number of circular triads occurring for these four cases is found to be

No. of Circular Triads.	Frequency.
0	2
1	2
2	4

We must multiply the frequency by three and by two to allow for similar symmetrical arrangements, and the final results are

No. of Circular Triads.	Frequency.
0	24
1	16
2	24
TOTAL	64

The principles of this method are clear enough and the work may be formalised by a number of conventions which we omit to save space. In common with many similar combinatorial problems, however, troubles arise from the sheer number of possibilities and the difficulty of ensuring that nothing is overlooked. Up to the present the distribution of d for n up to and including 7 is known. The frequencies and probabilities are given in Table 16.10.

Paired Comparisons for m *Observers: Coefficient of Agreement*

16.45. We now consider the investigation of similarities of judgments for m observers. Suppose that in a table of the form of Table 16.9 we enter a unit in the cell in row X and column Y whenever $X \to Y$ and count the units in each cell. A cell may then contain any number from 0 to m. If the observers are in complete agreement there will be $\binom{n}{2}$ cells containing the number m, the remaining $\binom{n}{2}$ cells being zero. The agreement may be complete even if there are inconsistences present.

Suppose that the cell in row X and column Y contains the number γ. Let

$$\Sigma = \Sigma \qquad (16.72)$$

the summation extending over the $n(n-1)$ cells of the table (the diagonal cells being ignored). Σ is then the sum of the number of agreements between pairs of judges. Put

$$u = \frac{2\Sigma}{[illegible]} - 1. \quad . \quad . \quad . \quad . \quad . \qquad (16.73)$$

TABLE 16.10

Paired Comparisons. Frequency (f) of Values of d and Probability (P) that Values will be Attained or Exceeded.

Value of d.	$n = 2$		$n = 3$		$n = 4$		$n = 5$		$n = 6$		$n = 7$	
	f	P	f	P	f	P	f	P	f	P	f	P
0	2	1·000	6	1·000	24	1·000	120	1·000	720	1·000	5,040	1·000
1			2	0·250	16	0·625	120	0·883	960	0·978	8,400	0·998
2					24	0·375	240	0·766	2,240	0·949	21,840	0·994
3							240	0·531	2,880	0·880	33,600	0·983
4							280	0·297	6,240	0·792	75,600	0·967
5							24	0·023	3,648	0·602	90,384	0·931
6									8,640	0·491	179,760	0·888
7									4,800	0·227	188,160	0·802
8									2,640	0·081	277,200	0·713
9											280,560	0·580
10											384,048	0·447
11											244,160	0·263
12											233,520	0·147
13											72,240	0·036
14											2,640	0·001
TOTAL	2	—	8	—	64	—	1,024	—	32,768	—	2,097,152	—

The maximum number of agreements, occurring if $\binom{n}{2}$ cells each contain m, is $\binom{n}{2}\binom{m}{2}$ and thus in the case of complete agreement, and only in this case, $u = 1$. The further we go from this case, as measured by agreements between pairs of observers, the smaller u becomes. The minimum number of agreements occurs when each cell contains $\frac{m}{2}$ if m is even or $\frac{(m \pm 1)}{2}$ if m is odd. That is, if m is even, the minimum number of agreements is

$$2\binom{\frac{m}{2}}{2}\binom{n}{2} = \tfrac{1}{4}m(m-2)\binom{n}{2},$$

and in this case

$$u = -\frac{1}{m-1}. \tag{16.74}$$

When m is odd the minimum value of u is found to be

$$u = -\frac{1}{m}. \tag{16.75}$$

16.46. We shall call u the Coefficient of Agreement. It is unity if and only if there is complete agreement in the comparisons. Its minimum value is not -1 except when $m = 2$. This, however, is to be expected in a measure of agreement, for there can be no such thing as complete disagreement among three or more observers in paired comparisons.

If observer P differs in certain comparisons from observers Q and R, the two latter must agree on those comparisons.

When $m = 2$, u reduces to

$$u = \frac{2\Sigma}{\binom{n}{2}} - 1 \qquad . \quad (16.76)$$

and Σ becomes twice the number of cases in which the two observers agree about a comparison. u is thus a generalisation of a coefficient τ. For general m, if the entries in the table were constrained to the ranking type, u would be the average intercorrelation τ between observers taken two at a time.

16.47. In discussing the significance of u it is desirable to know whether the set of preferences which give rise to it could have arisen by chance if the preferences had been assigned at random with respect to the quality under consideration. The procedure which first suggests itself is a generalisation of the method used for the case of m rankings. That is to say, we sum the entries in the rows of the table and consider the variance of these entries. If the preferences are allotted at random we expect to find about equal numbers given to each object, and the variance will be low; in other cases it will be higher.

The difficulty about this suggestion is that it has not been found possible to ascertain the distribution of the variance in the $2^{m\binom{n}{2}}$ possible sets of preferences. The case $m = 1$, corresponding to the distribution of d for inconsistences, is difficult enough to solve. For higher values of m no distributions are known except in trivial cases.

A test can, however, be devised by using the coefficient u. Consider one cell in the table in row X and column Y and let it contain the number γ. Then the corresponding cell in row Y and column X will contain $m - \gamma$. Thus these two contribute to Σ the amount $\binom{\gamma}{2} + \binom{m-\gamma}{2}$.

Now, of the total ways in which the units can be distributed in the first cell there will be $\binom{m}{\gamma}$ in which γ units occur. Consequently the distribution of Σ in the cell and the corresponding cell is given by the expression

$$f = t^{\binom{m}{2}} + m\, t^{\binom{m-1}{2}} + \quad t^{\binom{m-1}{2}+\binom{2}{2}} + \quad + \; m \quad t^{\binom{m-\gamma}{2}+\binom{\gamma}{2}} + \quad + t^{\binom{m}{2}} \quad (16.77)$$

and since the distribution in other pairs of cells is independent if the preferences are allotted at random the distribution of Σ for the whole table is given by

$$D(\Sigma) = f^N \qquad . \quad (16.78)$$

where $N = \binom{n}{2}$.

16.48. The distributions have been worked out for the following values of m and n: $m = 3$, $n = 2$ to 8; $m = 4$, $n = 2$ to 6; $m = 5$, $n = 2$ to 5; $m = 6$, $n = 2$ to 4. Tables 16.11 to 16.14 give the probabilities based on these distributions, i.e. the probabilities that a given value of Σ will be attained or exceeded.

For constant n the distribution tends to the Type III form as m tends to infinity. In fact, for a single pair of related cells the variate-value corresponding to a frequency $\binom{m}{\gamma}$ is $\binom{m-\gamma}{2} + \binom{\gamma}{2}$, which is a quadratic in γ. Were the variate-value a linear function of γ the distribution for the single cell would tend to normality in accordance with the well-known property of the binomial. The case of the quadratic value corresponds to a transformation of the variate of the type $x^2 = y$, and the transform of the normal form $\exp(-x^2)\,dx$ becomes the Type III form $\exp(-y)\,y^{-\frac{1}{2}}\,dy$. Since the N cells are independent and the sum of variates in the same Type III form is also distributed in that form, it follows that Σ is in the limit distributed as $\exp(-\Sigma)\,\Sigma^{\frac{N}{2}-1}\,d\Sigma$ except perhaps for some constants. Thus Σ or some multiple of it is distributed as χ^2.

For constant m the distribution tends to normality with increasing n.

TABLE 16.11

Agreement in Paired Comparisons. The Probability P that a Value of Σ will be Attained or Exceeded, for $m = 3$, $n = 2$ to 8.

$n = 2$		$n = 3$		$n = 4$		$n = 5$		$n = 6$		$n = 7$		$n = 8$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
1	1·000	3	1·000	6	1·000	10	1·000	15	1·000	21	1·000	28	1·000
3	0·250	5	0·578	8	0·822	12	0·944	17	0·987	23	0·998	30	1·000
		7	0·156	10	0·466	14	0·756	19	0·920	25	0·981	32	0·997
		9	0·016	12	0·169	16	0·474	21	0·764	27	0·925	34	0·983
				14	0·038	18	0·224	23	0·539	29	0·808	36	0·945
				16	0·0046	20	0·078	25	0·314	31	0·633	38	0·865
				18	$0{\cdot}0^{3}24$	22	0·020	27	0·148	33	0·433	40	0·736
						24	0·0035	29	0·057	35	0·256	42	0·572
						26	$0{\cdot}0^{3}42$	31	0·017	37	0·130	44	0·400
						28	$0{\cdot}0^{4}30$	33	0·0042	39	0·056	46	0·250
						30	$0{\cdot}0^{6}95$	35	$0{\cdot}0^{3}79$	41	0·021	48	0·138
								37	$0{\cdot}0^{3}12$	43	0·0064	50	0·068
								39	$0{\cdot}0^{4}12$	45	0·0017	52	0·029
								41	$0{\cdot}0^{6}92$	47	$0{\cdot}0^{3}37$	54	0·011
								43	$0{\cdot}0^{7}43$	49	$0{\cdot}0^{4}68$	56	0·0038
								45	$0{\cdot}0^{9}93$	51	$0{\cdot}0^{4}10$	58	0·0011
										53	$0{\cdot}0^{5}12$	60	$0{\cdot}0^{3}29$
										55	$0{\cdot}0^{6}12$	62	$0{\cdot}0^{4}66$
										57	$0{\cdot}0^{8}86$	64	$0{\cdot}0^{4}13$
										59	$0{\cdot}0^{9}44$	66	$0{\cdot}0^{5}22$
										61	$0{\cdot}0^{10}15$	68	$0{\cdot}0^{6}32$
										63	$0{\cdot}0^{12}23$	70	$0{\cdot}0^{7}40$
												72	$0{\cdot}0^{8}42$
												74	$0{\cdot}0^{9}36$
												76	$0{\cdot}0^{10}24$
												78	$0{\cdot}0^{11}13$
												80	$0{\cdot}0^{13}48$
												82	$0{\cdot}0^{14}12$
												84	$0{\cdot}0^{16}14$

TABLE 16.12

Agreement in Paired Comparisons. The Probability P that a Value of Σ will be Attained or Exceeded, for $m = 4$ and $n = 2$ to 6 (for $n = 6$ only Values beyond the 1 per cent. Point are given).

$n = 2$		$n = 3$		$n = 4$		$n = 5$		$n = 5$		$n = 6$		$n = 6$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
2	1·000	6	1·000	12	1·000	20	1·000	42	0·0048	57	0·014	79	$0{\cdot}0^{8}42$
3	0·625	7	0·947	13	0·997	21	1·000	43	0·0030	58	0·0092	80	$0{\cdot}0^{8}28$
6	0·125	8	0·736	14	0·975	22	0·999	44	0·0017	59	0·0058	81	$0{\cdot}0^{9}98$
		9	0·455	15	0·901	23	0·995	45	$0{\cdot}0^{3}73$	60	0·0037	82	$0{\cdot}0^{9}15$
		10	0·330	16	0·769	24	0·979	46	$0{\cdot}0^{3}41$	61	0·0022	83	$0{\cdot}0^{9}12$
		11	0·277	17	0·632	25	0·942	47	$0{\cdot}0^{3}24$	62	0·0013	84	$0{\cdot}0^{10}51$
		12	0·137	18	0·524	26	0·882	48	$0{\cdot}0^{4}90$	63	$0{\cdot}0^{3}76$	86	$0{\cdot}0^{11}30$
		14	0·043	19	0·410	27	0·805	49	$0{\cdot}0^{4}37$	64	$0{\cdot}0^{3}44$	87	$0{\cdot}0^{11}17$
		15	0·025	20	0·278	28	0·719	50	$0{\cdot}0^{4}25$	65	$0{\cdot}0^{3}23$	90	$0{\cdot}0^{13}28$
		18	0·0020	21	0·185	29	0·621	51	$0{\cdot}0^{5}93$	66	$0{\cdot}0^{3}13$		
				22	0·137	30	0·514	52	$0{\cdot}0^{5}21$	67	$0{\cdot}0^{4}72$		
				23	0·088	31	0·413	53	$0{\cdot}0^{5}17$	68	$0{\cdot}0^{4}36$		
				24	0·044	32	0·327	54	$0{\cdot}0^{6}74$	69	$0{\cdot}0^{4}18$		
				25	0·027	33	0·249	56	$0{\cdot}0^{7}66$	70	$0{\cdot}0^{5}97$		
				26	0·019	34	0·179	57	$0{\cdot}0^{7}38$	71	$0{\cdot}0^{5}47$		
				27	0·0079	35	0·127	60	$0{\cdot}0^{9}93$	72	$0{\cdot}0^{5}20$		
				28	0·0030	36	0·090			73	$0{\cdot}0^{5}10$		
				29	0·0025	37	0·060			74	$0{\cdot}0^{6}51$		
				30	0·0011	38	0·038			75	$0{\cdot}0^{6}18$		
				32	$0{\cdot}0^{3}16$	39	0·024			76	$0{\cdot}0^{7}78$		
				33	$0{\cdot}0^{4}95$	40	0·016			77	$0{\cdot}0^{7}44$		
				36	$0{\cdot}0^{5}38$	41	0·0088			78	$0{\cdot}0^{7}15$		

TABLE 16.13

Agreement in Paired Comparisons. The Probability P that a Value of Σ will be Attained or Exceeded, for $m = 5$ and $n = 2$ to 5.

$n = 2$		$n = 3$		$n = 4$		$n = 5$		$n = 5$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
4	1·000	12	1·000	24	1·000	40	1·000	76	$0{\cdot}0^{4}50$
6	0·375	14	0·756	26	0·940	42	0·991	78	$0{\cdot}0^{4}16$
10	0·063	16	0·390	28	0·762	44	0·945	80	$0{\cdot}0^{5}50$
		18	0·207	30	0·538	46	0·843	82	$0{\cdot}0^{5}15$
		20	0·103	32	0·353	48	0·698	84	$0{\cdot}0^{6}39$
		22	0·030	34	0·208	50	0·537	86	$0{\cdot}0^{6}10$
		24	0·011	36	0·107	52	0·384	88	$0{\cdot}0^{7}23$
		26	0·0039	38	0·053	54	0·254	90	$0{\cdot}0^{8}53$
		30	$0{\cdot}0^{3}24$	40	0·024	56	0·158	92	$0{\cdot}0^{8}12$
				42	0·0093	58	0·092	94	$0{\cdot}0^{9}14$
				44	0·0036	60	0·050	96	$0{\cdot}0^{10}46$
				46	0·0012	62	0·026	100	$0{\cdot}0^{12}91$
				48	$0{\cdot}0^{3}36$	64	0·012		
				50	$0{\cdot}0^{3}12$	66	0·0057		
				52	$0{\cdot}0^{4}28$	68	0·0025		
				54	$0{\cdot}0^{5}54$	70	0·0010		
				56	$0{\cdot}0^{5}18$	72	$0{\cdot}0^{3}39$		
				60	$0{\cdot}0^{7}60$	74	$0{\cdot}0^{3}14$		

TABLE 16.14

Agreement in Paired Comparisons. The Probability P that a Value of Σ will be Attained or Exceeded, for m = 6 and n = 2 to 4.

$n = 2$		$n = 3$		$n = 4$		$n = 4$		$n = 4$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
6	1·000	18	1·000	36	1·000	55	0·043	74	0·0^{4}12
7	0·688	19	0·969	37	0·999	56	0·029	75	0·0^{5}89
10	0·219	20	0·832	38	0·991	57	0·020	76	0·0^{5}49
15	0·031	21	0·626	39	0·959	58	0·016	77	0·0^{5}32
		22	0·523	40	0·896	59	0·011	80	0·0^{6}68
		23	0·468	41	0·822	60	0·0072	81	0·0^{6}17
		24	0·303	42	0·755	61	0·0049	82	0·0^{6}12
		26	0·180	43	0·669	62	0·0034	85	0·0^{7}34
		27	0·147	44	0·556	63	0·0025	90	0·0^{8}93
		28	0·088	45	0·466	64	0·0016		
		29	0·061	46	0·409	65	0·0^{3}83		
		30	0·040	47	0·337	66	0·0^{3}66		
		31	0·034	48	0·257	67	0·0^{3}48		
		32	0·023	49	0·209	68	0·0^{3}26		
		35	0·0062	50	0·175	69	0·0^{3}16		
		36	0·0029	51	0·133	70	0·0^{4}86		
		37	0·0020	52	0·097	71	0·0^{4}68		
		40	0·0^{3}58	53	0·073	72	0·0^{4}48		
		45	0·0^{4}31	54	0·057	73	0·0^{4}16		

16.49. The first of these results suggests that the Type III distribution will provide an approximation to the distribution (16.78) when m is moderately large. We proceed to find the first four moments of (16.78).

It is sufficient to find the first four moments of (16.77), those of (16.78) being obtainable therefrom in virtue of the relationships which connect cumulants of independent distributions.

The rth moment of (16.77) about the origin is given by

$$2^m\mu_r' = \left[\left(t\frac{\partial}{\partial t}\right)^r f\right]_{t=1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (16.79)$$

since 2^m is the total frequency. Thus we have

$$2^m\mu_1' = \sum_{r=0}^{m}\binom{m}{r}\left\{r^2 - mr + \frac{m^2-m}{2}\right\} = 2^m\binom{m}{2} + \Sigma\binom{m}{r}(r^2 - mr) \quad . \quad (16.80)$$

Sums such as $\Sigma\binom{m}{r}r^p$ can be obtained by operating on the binomial $(1+x)^m$ p times by $x\frac{\partial}{\partial x}$, e.g. we find

$$\Sigma\left\{\binom{m}{r}r\right\} = 2^m\frac{m}{2},$$

$$\Sigma\binom{m}{r}r^2 = 2^m\left\{\frac{m}{2} + \frac{1}{2}\binom{m}{2}\right\},$$

and hence, substituting in (16.80),

$$\mu_1' \quad \overline{2}\backslash 2$$

Thus the mean of the distribution (16.78) is given by

$$\mu_1' = \tfrac{1}{2}N(\qquad (16.81)$$

In a similar way we find

$$\left.\begin{aligned} \mu_2 &= \tfrac{1}{4}N \\ \mu_3 &= \tfrac{3}{4}N\binom{m}{3}, \\ \mu_4 &= N\binom{m}{2}\left\{\frac{3m^2 - 15m + 17}{8} + \tfrac{3}{32}N(m^2 - m)\right\} \end{aligned}\right\} \qquad (16.82)$$

These are the moments of Σ. Those of u are obtained by dividing by an appropriate power of $N\binom{m}{2}$ and it may be noted in particular that the mean of u is zero.

16.50. The first four moments of the Type III distribution

$$dF = ke^{-px}\,x^{q-1}\,dx$$

are

$$\frac{q}{p},\ \frac{q}{p^2},\ \frac{2q}{p^3},\ \frac{3q(q+2)}{p^4}$$

Equating the second and third moments to those given by (16.82) we find

$$\left.\begin{aligned} q &= \frac{Nm(m-1)}{2(m-2)^2}, \\ p &= \frac{2}{m-2}. \end{aligned}\right\} \qquad (16.83)$$

To make the first moments correspond we move the origin of the Σ-distribution a distance $\tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}$ to the right. We thus reach the approximation to the Σ-distribution, coinciding in the first three moments,

$$dF = ke^{-\frac{2x}{m-2}}\,x^{\frac{Nm(m-1)}{2(m-1)^2}-1}\,dx,$$

where

$$x = \Sigma - \tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}$$

or, transforming to the more usual χ^2 form by putting $\chi^2 = \frac{4x}{m-2}$, we find that

$$\tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}\Big\}\frac{}{m-2} \qquad (16.84)$$

is distributed as χ^2 with

$$\nu = \frac{Nm(m-1)}{(m \quad 2)^2} \qquad (16.85)$$

degrees of freedom.

The fourth moments of Σ and the χ^2 approximation differ by terms of order N^{-1} and m^{-1} compared with their absolute values.

16.51. It only remains to be seen how large m and n must be for this to provide a satisfactory approximation.

Consider first the distributions for $m = 3$. When $n = 8$, $N = 28$, we have, for the approximation, 4Σ distributed with 168 degrees of freedom. From Table 16.11 we see that for $\Sigma = 54$, $P = 0{\cdot}011$ and for $\Sigma = 58$, $P = 0{\cdot}0011$. Applying a continuity correction by deducting unity from Σ we find for the χ^2 approximation with $\chi^2 = 4 \times 53$, $\nu = 168$, $P = 0{\cdot}011$, and with $\chi^2 = 4 \times 57$, $P = 0{\cdot}00114$. The correspondence is very close, in spite of the low value of m.

For $m = 4$, $n = 5$, $N = 10$, the approximation gives $2\Sigma - 30$ distributed with 30 degrees of freedom. For $\Sigma = 40$ and 41, this gives, with continuity corrections of 0·5, half the variate-interval, $\chi^2 = 49$ and 51, $\nu = 30$. From the diagram at the end it is seen that these values lie one on either side of the 1 per cent. value; and this is in accordance with the exact values of P, which are seen from Table 16.12 to be 0·016 and 0·0088. Similarly we find that the values of Σ, 37 and 38, lie on either side of the 5 per cent. level, which is again in accordance with the exact values, $P = 0{\cdot}060$ and 0·038.

For $m = 6$, $n = 4$, $N = 6$, the approximation gives $\Sigma - 33{\cdot}75$ distributed with 11·25 degrees of freedom. For $\Sigma = 59$ and 60 the corresponding χ^2 values are seen to lie on either side of the 1 per cent. point, which accords with the exact value of Table 16.14.

We conclude that the χ^2 approximation provides an adequate test of significance for the values of m and n outside the range for which Tables 16.13 and 16.14 give exact values.

Example 16.7

A class of boys (ages 11 to 13 inclusive) were asked to state their preferences with respect to certain school subjects. Each child was given a sheet on which were written the possible pairs of subjects and asked to underline the one preferred in each case. The results were as follows:

21 boys, 13 school subjects. The preferences are shown in Table 16.15, which is in the form described in **16.39**; e.g. there were 18 boys who preferred Art to Religion.

TABLE 16.15

Preferences of 21 Boys in 13 Subjects.

	1	2	3	4	5	6	7	8	9	10	11	12	13	TOTALS
1. Woodwork	—	14	20	15	15	16	16	18	18	18	20	21	20	211
2. Gymnastics	7	—	14	12	13	18	14	16	16	20	16	18	19	183
3. Art	1	7	—	10	14	10	16	18	16	16	17	16	19	160
4. Science	6	9	11	—	11	12	15	14	13	13	17	17	16	154
5. History	6	8	7	10	—	14	11	12	14	15	13	14	16	140
6. Geography	5	3	11	9	7	—	14	14	13	13	16	15	17	137
7. Arithmetic	5	7	5	6	10	7	—	9	11	13	15	13	15	116
8. Religion	3	5	3	7	9	7	12	—	12	14	14	16	14	116
9. English Literature	3	5	5	8	7	8	10	9	—	10	13	13	15	106
10. Commercial subjects	3	1	5	8	6	8	8	7	11	—	10	10	14	91
11. Algebra	1	5	4	4	8	5	6	7	8	11	—	10	13	82
12. English Grammar	0	3	5	4	7	6	8	5	8	11	11	—	13	81
13. Geometry	1	2	2	5	5	4	6	7	6	7	8	8	—	61
													TOTAL	1638

The calculation of Σ for this table, in which the objects are arranged in order of total number of preferences, may be shortened by noting that Σ, as given by equation (16.72), may be transformed into the form

$$\Sigma = \Sigma(\gamma^2) - m\Sigma(\gamma) + \binom{m}{2}\binom{n}{2},$$

where the summation now takes place over the half of the table below the diagonal. Since the numbers in this half are smaller than those in the other half there is a considerable saving in arithmetic.

We find $$\Sigma = 9718$$

and hence $$u = \frac{2 \times 9718}{\binom{21}{2}\binom{13}{2}} - 1 = 0{\cdot}186.$$

There is thus a certain amount of agreement among the children, indicated by the positive value of u. Is this significant?

We note first of all that this distribution of preferences could not have arisen by chance to any acceptable degree of probability. In fact, $\chi^2 = 412{\cdot}4$ (equation 16.84)) and $\nu = 90{\cdot}7$. The large value of ν justifies the use of the normal approximation to the χ^2-distribution and we find $\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = 15{\cdot}3$, a very improbable result on the hypothesis of a random allocation of preferences.

The distribution of circular triads was as follows :—

No. of Triads.	Frequency.	No. of Triads.	Frequency.
0	1	12	1
1	1	17	3
4	5	21	1
6	2	25	1
7	2	29	1
8	1	39	1
10	1		
		TOTAL	21

The total number of circular triads was 242 with a mean of 11·5. Only one boy was entirely consistent. On the other hand, for $n = 13$ the maximum number of circular triads is 91, with a mathematical expectation of 71·5. It is thus clear that, except perhaps for one boy, we cannot suppose that any boy allotted preferences at random. We are again led to conclude that the boys are genuinely capable of making distinctions, and that consistently on the whole. Half the boys have coefficients of consistence ζ greater than 0·92.

We conclude that the boys can make preferences and that in their view the subjects are sufficiently different to enable a reasonably consistent set of preferences to be made. So far as these data are concerned there would be no objection to the assumption that a *scale* of preferences can be set up. With this in mind, we can say that the value of u indicates a certain amount of agreement, though not a strong one, between the boys as to which subjects they prefer.

NOTES AND REFERENCES

Spearman has suggested another coefficient of rank correlation, viz.

$$R = 1 - \frac{3\Sigma\,|\,d\,|}{n^2 - 1},$$

but this "footrule" is unreliable as a measure of dependence—it cannot, for example, attain -1. For earlier work on rank correlation see Spearman (1904, 1906), K. Pearson (1907) and "Student" (1921). The distribution of ρ in the case of independence was given by Kendall and others (1939). Pitman (1937) had previously suggested that it could be approximately represented by the B-distribution.

The coefficient τ was suggested by Kendall in 1938. In practice ρ is probably more convenient. It is, however, remarkable that τ is unique among correlation coefficients in depending only on linear processes, so that machines may be constructed to calculate it. Furthermore, τ can be adapted to give partial rank correlation coefficients (Kendall, 1942).

The problem of m rankings was considered by Friedman in 1937 and by Babington Smith and Kendall and by Wallis in 1939. Friedman (1940) has reviewed this work and provided some useful tables based on the Type I approximative distribution. Wallis has pointed out that the coefficient W is the ranking analogue of the correlation ratio. Kelley (*Statistical Method*) had considered ρ_{av} as a measure of concordance in rankings.

REFERENCES

Eells, W. C. (1929), "Formulas for probable errors of coefficients of correlation," *J. Amer. Statist. Ass.*, **24**, 170.

Friedman, M. (1937), "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Jour. Amer. Statist. Ass.*, **32**, 675.

—— (1940), "A comparison of alternative tests of significance for the problem of m rankings," *Ann. Math. Statist.*, **11**, 86.

Hotelling, H., and Pabst, M. R. (1936), "Rank correlation and tests of significance involving no assumption of normality," *Ann. Math. Statist.*, **7**, 29.

Kendall, M. G. (1938), "A new measure of rank correlation," *Biometrika*, **30**, 81.

——, M. G., Kendall, S. F. H., and Babington Smith, B. (1939), "The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times," *Biometrika*, **30**, 251.

—— and Babington Smith, B. (1939), "The problem of m rankings," *Ann. Math. Statist.*, **10**, 275.

—— and Babington Smith, B. (1940), "On the method of paired comparisons," *Biometrika*, **31**, 324.

—— (1942), "Partial Rank Correlation," *Biometrika*, **32**, 277.

Pearson, K. (1907), "On further methods of determining correlation," *Drapers Co. Memoirs, Biometric Series IV*, London, Dulau & Co.

Pitman, E. J. G. (1937), "Significance tests which may be applied to samples from any populations; Part II. The correlation coefficient test," *J. Roy. Statist. Supp.*, **4**, 225, and (1938) "Part III. The analysis of variance," *Biometrika*, **29**, 322.

Spearman, C. (1904), "The proof and measurement of association between two things, *Amer. J. Psychol.*, **15**, 88.

—— (1906), "A footrule for measuring correlation," *Brit. Jour. Psychol.*, **2**, 89.

"Student" (1921), "An experimental determination of the probable error of Dr. Spearman's correlation coefficients," *Biometrika*, **13**, 263.

Wallis, W. A. (1939), "The correlation ratio for ranked data," *Jour. Amer. Statist. Ass.*, **34**, 533.

EXERCISES

16.1. Show that the coefficients of rank correlation ρ between the natural order 1, . . . 10 and the following rankings are $-0{\cdot}37$ and $+0{\cdot}45$ respectively.

$$7,\ 10,\ 4,\ 1,\ 6,\ 8,\ 9,\ 5,\ 2,\ 3\,;$$
$$10,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9.$$

Show that the corresponding values of τ are $-0{\cdot}24$ and $+0{\cdot}60$.

16.2. Defining

$$\chi_r^2 = m(n-1)\ W$$

show that approximately χ_r^2 is distributed as χ^2 in the Type III form with $\nu = n-1$ degrees of freedom.

(Friedman, 1937.)

16.3. Show that W is the ratio of the sum of squares between columns and the total sum of squares (the rankings being regarded as arrayed one below the other) and hence that W is the square of the correlation ratio η_{yx}^2 for such an array (the ranks being regarded as variate-values). The "sum of squares between columns" means the sum of squares of deviations of column means from their mean.

(Wallis, 1939.)

16.4. Show that Spearman's "footrule"

$$R = 1 - \frac{3\Sigma\,|\,d\,|}{n^2-1}$$

can attain, but not exceed, the value 1, and can be as small as, but not smaller than, $-\frac{1}{2}$.

16.5. Verify formula (16.63).

16.6. The following table shows the preferences of 25 girls in 11 school subjects.

	1	2	3	4	5	6	7	8	9	10	11	Totals
1. Gymnastics	—	10	19	17	20	17	21	21	21	18	22	186
2. Science	15	—	12	15	17	15	21	19	18	16	17	165
3. Art	6	13	—	16	16	18	10	17	16	19	16	147
4. Domestic Science	8	10	9	—	16	11	13	15	14	11	14	121
5. History	5	8	9	9	—	14	18	12	13	15	18	121
6. Arithmetic	8	10	7	14	11	—	12	13	12	16	18	121
7. Geography	4	4	15	12	7	13	—	14	15	14	14	112
8. English Literature	4	6	8	10	13	12	11	—	14	13	14	105
9. Religion	4	7	9	11	12	13	10	11	—	11	17	105
10. Algebra	7	9	6	14	10	9	11	12	14	—	12	104
11. English Grammar	3	8	9	11	7	7	11	11	8	13	—	88
											Total	1375

Show that the coefficient of agreement u is $0{\cdot}082$; that this is significant; but that the girls are less alike in preferences than the boys of Example 16.7.

APPENDIX TABLE 1

Normal Distribution. Frequency Function of the Normal Distribution at every Tenth of the Standard Deviation, with First and Second Differences. The value of the central ordinate at zero is $1/\sqrt{2\pi}$.

$\frac{x}{\sigma}$.	y.	$\Delta^1(-)$.	Δ^2.	$\frac{x}{\sigma}$.	y.	$\Delta^1(-)$.	Δ^2.
0·0	0·39894	199	− 392	2·5	0·01753	395	+ 79
0·1	0·39695	591	− 374	2·6	0·01358	316	+ 66
0·2	0·39104	965	− 347	2·7	0·01042	250	+ 53
0·3	0·38139	1312	− 308	2·8	0·00792	197	+ 45
0·4	0·36827	1620	− 265	2·9	0·00595	152	+ 36
0·5	0·35207	1885	− 212	3·0	0·00443	116	+ 27
0·6	0·33322	2097	− 159	3·1	0·00327	89	+ 23
0·7	0·31225	2256	− 104	3·2	0·00238	66	+ 17
0·8	0·28969	2360	− 52	3·3	0·00172	49	+ 13
0·9	0·26609	2412	0	3·4	0·00123	36	+ 10
1·0	0·24197	2412	+ 46	3·5	0·00087	26	+ 7
1·1	0·21785	2366	+ 84	3·6	0·00061	19	+ 6
1·2	0·19419	2282	+ 118	3·7	0·00042	13	+ 4
1·3	0·17137	2164	+ 143	3·8	0·00029	9	+ 2
1·4	0·14973	2021	+ 161	3·9	0·00020	7	+ 3
1·5	0·12952	1860	+ 173	4·0	0·00013	4	—
1·6	0·11092	1687	+ 177	4·1	0·00009	3	—
1·7	0·09405	1510	+ 177	4·2	0·00006	2	—
1·8	0·07895	1333	+ 170	4·3	0·00004	2	—
1·9	0·06562	1163	+ 162	4·4	0·00002	—	—
2·0	0·05399	1001	+ 150	4·5	0·00002	—	—
2·1	0·04398	851	+ 137	4·6	0·00001	—	—
2·2	0·03547	714	+ 120	4·7	0·00001	—	—
2·3	0·02833	594	+ 108	4·8	0·00000	—	—
2·4	0·02239	486	+ 91				

Precision of Interpolation.—Owing to the magnitude of the second differences, simple interpolation near the beginning of the table may give an error up to 5 in the fourth place; the use of second differences will bring this down to 1 or 2 in the last place, third differences being small. Where third differences are greatest, in the neighbourhood of $x/\sigma = 0{\cdot}6$, the error may be as large as 3 in the last place unless the third difference is used.

APPENDIX TABLE 2

Normal Distribution. The Distribution Function F of the Normal Distribution, tabulated at every Tenth of the Standard Deviation, with First and Second Differences.

$\frac{x}{\sigma}$.	F.	$\Delta^1(+)$.	$\Delta^2(-)$.	$\frac{x}{\sigma}$.	F.	$\Delta^1(+)$.	$\Delta^2(-)$.
0·0	0·50000	3983	40	2·5	0·99379	155	36
0·1	0·53983	3943	78	2·6	0·99534	119	28
0·2	0·57926	3865	114	2·7	0·99653	91	22
0·3	0·61791	3751	147	2·8	0·99744	69	17
0·4	0·65542	3604	175	2·9	0·99813	52	14
0·5	0·69146	3429	200	3·0	0·99865	38	10
0·6	0·72575	3229	219	3·1	0·99903	28	7
0·7	0·75804	3010	230	3·2	0·99931	21	7
0·8	0·78814	2780	240	3·3	0·99952	14	3
0·9	0·81594	2540	241	3·4	0·99966	11	4
1·0	0·84134	2299	239	3·5	0·99977	7	—
1·1	0·86433	2060	233	3·6	0·99984	5	—
1·2	0·88493	1827	223	3·7	0·99989	4	—
1·3	0·90320	1604	209	3·8	0·99993	2	—
1·4	0·91924	1395	194	3·9	0·99995	2	—
1·5	0·93319	1201	178	4·0	0·99997	1	—
1·6	0·94520	1023	159	4·1	0·99998	1	—
1·7	0·95543	864	143	4·2	0·99999	—	—
1·8	0·96407	721	124	4·3	0·99999	—	—
1·9	0·97128	597	108	4·4	0·99999	—	—
2·0	0·97725	489	93				
2·1	0·98214	396	78				
2·2	0·98610	318	66				
2·3	0·98928	252	53				
2·4	0·99180	199	44				

F attains the exact value 0·99999 between 4·26 and 4·27.

Precision of Interpolation.—Simple interpolation may lead to an error of 3 or 4 at most in the fourth place of decimals in the region where second differences are large; the use of the second difference will bring this down to 2 or 3 in the last place, the largest errors tending to occur at the beginning of the table, where the third difference may be used if the greatest possible precision is desired.

t-Table. The Distribution Function of $y = \dfrac{y_0}{\left(1 + \dfrac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}$ *for Values of t*

(Condensed to three figures from the four-figure

t.	ν = 1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
0	0·500	0·500	0·500	0·500	0·500	0·500	0·500	0·500	0·500	0·500
0·1	0·532	0·535	0·537	0·537	0·538	0·538	0·538	0·539	0·539	0·539
0·2	0·563	0·570	0·573	0·574	0·575	0·576	0·576	0·577	0·577	0·577
0·3	0·593	0·604	0·608	0·610	0·612	0·613	0·614	0·614	0·614^5	0·615
0·4	0·621	0·636	0·642	0·645	0·647	0·648^5	0·649^5	0·650	0·651	0·651
0·5	0·648	0·667	0·674	0·678	0·681	0·683	0·684	0·685	0·685^5	0·686
0·6	0·672	0·695	0·705	0·710	0·713	0·715	0·716	0·717	0·718	0·719
0·7	0·694	0·722	0·733	0·739	0·742	0·745	0·747	0·748	0·749	0·750
0·8	0·715	0·746	0·759	0·766	0·770	0·773	0·775	0·777	0·772	0·779
0·9	0·733	0·768	0·783	0·790^5	0·795	0·799	0·801	0·803	0·804	0·805
1·0	0·750	0·789	0·804^5	0·813	0·818	0·822	0·825	0·827	0·828	0·830
1·1	0·765	0·807	0·824	0·833^5	0·839	0·843	0·846	0·848	0·850	0·851
1·2	0·779	0·823^5	0·842	0·852	0·858	0·862	0·865	0·868	0·870	0·871
1·3	0·791	0·838	0·858	0·868	0·875	0·879	0·883	0·885	0·887	0·889
1·4	0·803	0·852	0·872	0·883	0·890	0·894^5	0·898	0·900^5	0·902^5	0·904
1·5	0·813	0·864	0·885	0·896	0·903	0·908	0·911	0·914	0·916	0·918
1·6	0·822	0·875	0·896	0·908	0·915	0·920	0·923	0·926	0·928	0·930
1·7	0·831	0·884	0·906	0·918	0·925	0·930	0·933^5	0·936	0·938	0·940
1·8	0·839	0·893	0·915	0·927	0·934	0·939	0·943	0·945	0·947	0·949
1·9	0·846	0·901	0·923	0·935	0·942	0·947	0·950	0·953	0·955	0·957
2·0	0·852	0·908	0·930	0·942	0·949	0·954	0·957	0·960	0·962	0·963
2·1	0·858^5	0·915	0·937	0·948	0·955	0·960	0·963	0·965^5	0·967	0·969
2·2	0·864	0·921	0·942	0·954	0·960^5	0·965	0·968	0·970^5	0·972	0·974
2·3	0·869^5	0·926	0·947^5	0·958^5	0·965	0·969	0·972^5	0·975	0·976^5	0·978
2·4	0·874	0·931	0·952	0·963	0·969	0·973	0·976	0·978	0·980	0·981
2·5	0·879	0·935	0·956	0·967	0·973	0·977	0·979^5	0·981^5	0·983	0·984
2·6	0·883	0·939	0·960	0·970	0·976	0·980	0·982	0·984	0·986	0·987
2·7	0·887	0·943	0·963	0·973	0·979	0·982	0·985	0·986^5	0·988	0·989
2·8	0·891	0·946	0·966	0·976	0·981	0·984	0·987	0·988	0·990	0·991
2·9	0·894	0·949	0·969	0·978	0·983	0·986	0·988^5	0·990	0·991	0·992
3·0	0·898	0·952	0·971	0·980	0·985	0·988	0·990	0·991^5	0·992^5	0·993
3·1	0·901	0·955	0·973	0·982	0·987	0·989	0·991	0·993	0·994	0·994
3·2	0·904	0·957	0·975	0·983^5	0·988	0·991	0·992^5	0·994	0·995	0·995
3·3	0·906	0·960	0·977	0·985	0·989	0·992	0·993	0·995	0·995	0·996
3·4	0·909	0·962	0·979	0·986	0·990	0·993	0·994	0·995	0·996	0·997
3·5	0·911	0·964	0·980	0·988	0·991	0·994	0·995	0·996	0·997	0·997
3·6	0·914	0·965	0·982	0·989	0·992	0·994	0·996	0·996^5	0·997	0·998
3·7	0·916	0·967	0·983	0·990	0·993	0·995	0·996	0·997	0·997^5	0·998
3·8	0·918	0·969	0·984	0·990	0·994	0·995^5	0·997	0·997	0·998	0·998
3·9	0·920	0·970	0·985	0·991	0·994	0·996	0·997	0·998	0·998	0·998^5
4·0	0·922	0·971	0·986	0·992	0·995	0·996	0·997	0·998	0·998	0·999
4·1	0·924	0·973	0·987	0·993	0·995	0·997	0·998	0·998	0·999	0·999
4·2	0·926	0·974	0·988	0·993	0·996	0·997	0·998	0·998^5	0·999	0·999
4·3	0·927	0·975	0·988	0·994	0·996	0·997^5	0·998	0·999	0·999	0·999
4·4	0·929	0·976	0·989	0·994	0·996^5	0·998	0·998	0·999	0·999	0·999
4·5	0·930	0·977	0·990	0·995	0·997	0·998	0·999	0·999	0·999	0·999
4·6	0·932	0·978	0·990	0·995	0·997	0·998	0·999	0·999	0·999	0·999^5
4·7	0·933	0·979	0·991	0·995	0·997	0·998	0·999	0·999	0·999	1·000
4·8	0·935	0·980	0·991	0·996	0·998	0·998^5	0·999	0·999	0·999^5	
4·9	0·936	0·980	0·992	0·996	0·998	0·999	0·999	0·999	1·000	
5·0	0·937	0·981	0·992	0·996	0·998	0·999	0·999	0·999^5		
5·1	0·938	0·982	0·993	0·996^5	0·998	0·999	0·999	0·999^5		
5·2	0·939^5	0·982^5	0·993	0·997	0·998	0·999	0·999	1·000		
5·3	0·941	0·983	0·993	0·997	0·998	0·999	0·999			
5·4	0·942	0·984	0·994	0·997	0·998^5	0·999	0·999^5			
5·5	0·943	0·984	0·994	0·997	0·999	0·999	0·999^5			
5·6	0·944	0·985	0·994	0·997^5	0·999	0·999	1·000			
5·7	0·945	0·985	0·995	0·998	0·999	0·999				
5·8	0·946	0·986	0·995	0·998	0·999	0·999				
5·9	0·947	0·986	0·995	0·998	0·999	0·999^5				
6·0	0·947	0·987	0·995	0·998	0·999	0·999^5				

TABLE 3

proceeding by Intervals of 0·1 from 0 to 6, and for Values of ν from 1 to 20.

tables by "Student" in *Metron*, **5**, 1925.)

t.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
0	0·500	0·500	0·500	0·500	0·500	0·500	0·500	0·500	0·500	0·500
0·1	0·539	0·539	0·539	0·539	0·539	0·539	0·539	0·539	0·539	0·539
0·2	0·577	0·578	0·578	0·578	0·578	0·578	0·578	0·578	0·578	0·578
0·3	0·615	0·615	0·615^5	0·616	0·616	0·616	0·616	0·616	0·616	0·616
0·4	0·652	0·652	0·652	0·652	0·653	0·653	0·653	0·653	0·653	0·653
0·5	0·686^5	0·687	0·687	0·688	0·688	0·688	0·688	0·688	0·689	0·689
0·6	0·720	0·720	0·721	0·721	0·721	0·721^5	0·722	0·722	0·722	0·722
0·7	0·751	0·751	0·752	0·752	0·753	0·753	0·753	0·754	0·754	0·754
0·8	0·780	0·780	0·781	0·781^5	0·782	0·782	0·783	0·783	0·783	0·783
0·9	0·806	0·807	0·808	0·808	0·809	0·809	0·810	0·810	0·810	0·811
1·0	0·831	0·831^5	0·832	0·833	0·833	0·834	0·834	0·835	0·835	0·835
1·1	0·853	0·853^5	0·854	0·855	0·856	0·856	0·857	0·857	0·857^5	0·858
1·2	0·872	0·873	0·874	0·875	0·876	0·876	0·877	0·877	0·878	0·878
1·3	0·890	0·891	0·892	0·893	0·893	0·894	0·894^5	0·895	0·895	0·896
1·4	0·905^5	0·907	0·907^5	0·908	0·909	0·910	0·910	0·911	0·911	0·912
1·5	0·919	0·920	0·921	0·922	0·923	0·923^5	0·924	0·924^5	0·925	0·925
1·6	0·931	0·932	0·933	0·934	0·935	0·935	0·936	0·936^5	0·937	0·937
1·7	0·941	0·943	0·943^5	0·944	0·945	0·946	0·946	0·947	0·947	0·948
1·8	0·950	0·951^5	0·952^5	0·953	0·954	0·955	0·955	0·956	0·956	0·956^5
1·9	0·958	0·959	0·960	0·961	0·962	0·962	0·963	0·963	0·964	0·964
2·0	0·965	0·966	0·967	0·967	0·968	0·969	0·969	0·970	0·970	0·970
2·1	0·970	0·971	0·972	0·973	0·973^5	0·974	0·974^5	0·975	0·975	0·976
2·2	0·975	0·976	0·977	0·977	0·978	0·979	0·979	0·979	0·980	0·980
2·3	0·979	0·980	0·981	0·981	0·982	0·982	0·983	0·983	0·983^5	0·984
2·4	0·982	0·983	0·984	0·985	0·985	0·985^5	0·986	0·986	0·987	0·987
2·5	0·985	0·986	0·987	0·987	0·988	0·988	0·988^5	0·989	0·989	0·989
2·6	0·988	0·988	0·989	0·989^5	0·990	0·990	0·991	0·991	0·991	0·991
2·7	0·990	0·990	0·991	0·991	0·992	0·992	0·992	0·993	0·993	0·993
2·8	0·991	0·992	0·992^5	0·993	0·993	0·994	0·994	0·994	0·994	0·994^5
2·9	0·993	0·993	0·994	0·994	0·994^5	0·994^5	0·995	0·995	0·995	0·996
3·0	0·994	0·994^5	0·995	0·995	0·995^5	0·996	0·996	0·996	0·996	0·996^5
3·1	0·995	0·995	0·996	0·996	0·996	0·997	0·997	0·997	0·997	0·997
3·2	0·996	0·996	0·996^5	0·997	0·997	0·997	0·997	0·997^5	0·998	0·998
3·3	0·996^5	0·997	0·997	0·997	0·998	0·998	0·998	0·998	0·998	0·998
3·4	0·997	0·997	0·998	0·998	0·998	0·998	0·998	0·998	0·998^5	0·999
3·5	0·997^5	0·998	0·998	0·998	0·998	0·998^5	0·999	0·999	0·999	0·999
3·6	0·998	0·998	0·998	0·999	0·999	0·999	0·999	0·999	0·999	0·999
3·7	0·998	0·998^5	0·999	0·999	0·999	0·999	0·999	0·999	0·999	0·999
3·8	0·998^5	0·999	0·999	0·999	0·999	0·999	0·999	0·999	0·999	0·999
3·9	0·999	0·999	0·999	0·999	0·999	0·999	0·999	0·999^5	0·999^5	1·000
4·0	0·999	0·999	0·999	0·999	0·999	0·999^5	0·999^5	1·000	1·000	
4·1	0·999	0·999	0·999	0·999^5	0·999^5	1·000	1·000			
4·2	0·999	0·999	0·999^5	1·000	1·000					
4·3	0·999	0·999^5	1·000							
4·4	0·999^5	1·000								
4·5	0·999^5									
4·6	1·000									

Note.—The methods by which "Student" calculated the *Metron* tables are explained in notes by him and R. A. Fisher in that journal, vol. 5, Part 3, 1925, pp. 18–24. The four figures of those values have been rounded up to three in the above table, except when the four-figure value concluded with a 5, in which case it is shown in full. In columns in which values greater than 0·9995 occur the first is written 1·000 and the remainder left blank.

APPENDIX TABLE 4

(Reprinted from Table VI of Prof. R. A. Fisher's *Statistical Methods for Research Workers*, Oliver and Boyd, Ltd., Edinburgh, by kind permission of the author and the publishers.)

5 Per Cent. Points of the Distribution of z.

Values of ν_2.	Values of ν_1. 1.	2.	3.	4.	5.	6.	8.	12.	24.	∞.
1	2·5421	2·6479	2·6870	2·7071	2·7194	2·7276	2·7380	2·7484	2·7588	2·7693
2	1·4592	1·4722	1·4765	1·4787	1·4800	1·4808	1·4819	1·4830	1·4840	1·4851
3	1·1577	1·1284	1·1137	1·1051	1·0994	1·0953	1·0899	1·0842	1·0781	1·0716
4	1·0212	0·9690	0·9429	0·9272	0·9168	0·9093	0·8993	0·8885	0·8767	0·8639
5	0·9441	0·8777	0·8441	0·8236	0·8097	0·7997	0·7862	0·7714	0·7550	0·7368
6	0·8948	0·8188	0·7798	0·7558	0·7394	0·7274	0·7112	0·6931	0·6729	0·6499
7	0·8606	0·7777	0·7347	0·7080	0·6896	0·6761	0·6576	0·6369	0·6134	0·5862
8	0·8355	0·7475	0·7014	0·6725	0·6525	0·6378	0·6175	0·5945	0·5682	0·5371
9	0·8163	0·7242	0·6757	0·6450	0·6238	0·6080	0·5862	0·5613	0·5324	0·4979
10	0·8012	0·7058	0·6553	0·6232	0·6009	0·5843	0·5611	0·5346	0·5035	0·4657
11	0·7889	0·6909	0·6387	0·6055	0·5822	0·5648	0·5406	0·5126	0·4795	0·4387
12	0·7788	0·6786	0·6250	0·5907	0·5666	0·5487	0·5234	0·4941	0·4592	0·4156
13	0·7703	0·6682	0·6134	0·5783	0·5535	0·5350	0·5089	0·4785	0·4419	0·3957
14	0·7630	0·6594	0·6036	0·5677	0·5423	0·5233	0·4964	0·4649	0·4269	0·3782
15	0·7568	0·6518	0·5950	0·5585	0·5326	0·5131	0·4855	0·4532	0·4138	0·3628
16	0·7514	0·6451	0·5876	0·5505	0·5241	0·5042	0·4760	0·4428	0·4022	0·3490
17	0·7466	0·6393	0·5811	0·5434	0·5166	0·4964	0·4676	0·4337	0·3919	0·3366
18	0·7424	0·6341	0·5753	0·5371	0·5099	0·4894	0·4602	0·4255	0·3827	0·3253
19	0·7386	0·6295	0·5701	0·5315	0·5040	0·4832	0·4535	0·4182	0·3743	0·3151
20	0·7352	0·6254	0·5654	0·5265	0·4986	0·4776	0·4474	0·4116	0·3668	0·3057
21	0·7322	0·6216	0·5612	0·5219	0·4938	0·4725	0·4420	0·4055	0·3599	0·2971
22	0·7294	0·6182	0·5574	0·5178	0·4894	0·4679	0·4370	0·4001	0·3536	0·2892
23	0·7269	0·6151	0·5540	0·5140	0·4854	0·4636	0·4325	0·3950	0·3478	0·2818
24	0·7246	0·6123	0·5508	0·5106	0·4817	0·4598	0·4283	0·3904	0·3425	0·2749
25	0·7225	0·6097	0·5478	0·5074	0·4783	0·4562	0·4244	0·3862	0·3376	0·2685
26	0·7205	0·6073	0·5451	0·5045	0·4752	0·4529	0·4209	0·3823	0·3330	0·2625
27	0·7187	0·6051	0·5427	0·5017	0·4723	0·4499	0·4176	0·3786	0·3287	0·2569
28	0·7171	0·6030	0·5403	0·4992	0·4696	0·4471	0·4146	0·3752	0·3248	0·2516
29	0·7155	0·6011	0·5382	0·4969	0·4671	0·4444	0·4117	0·3720	0·3211	0·2466
30	0·7141	0·5994	0·5362	0·4947	0·4648	0·4420	0·4090	0·3691	0·3176	0·2419
60	0·6933	0·5738	0·5073	0·4632	0·4311	0·4064	0·3702	0·3255	0·2654	0·1644
∞	0·6729	0·5486	0·4787	0·4319	0·3974	0·3706	0·3309	0·2804	0·2085	0

APPENDIX TABLE 5

(Reprinted from Table VI of Prof. R. A. Fisher's *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh, by kind permission of the author and the publishers.)

1 Per Cent. Points of the Distribution of z.

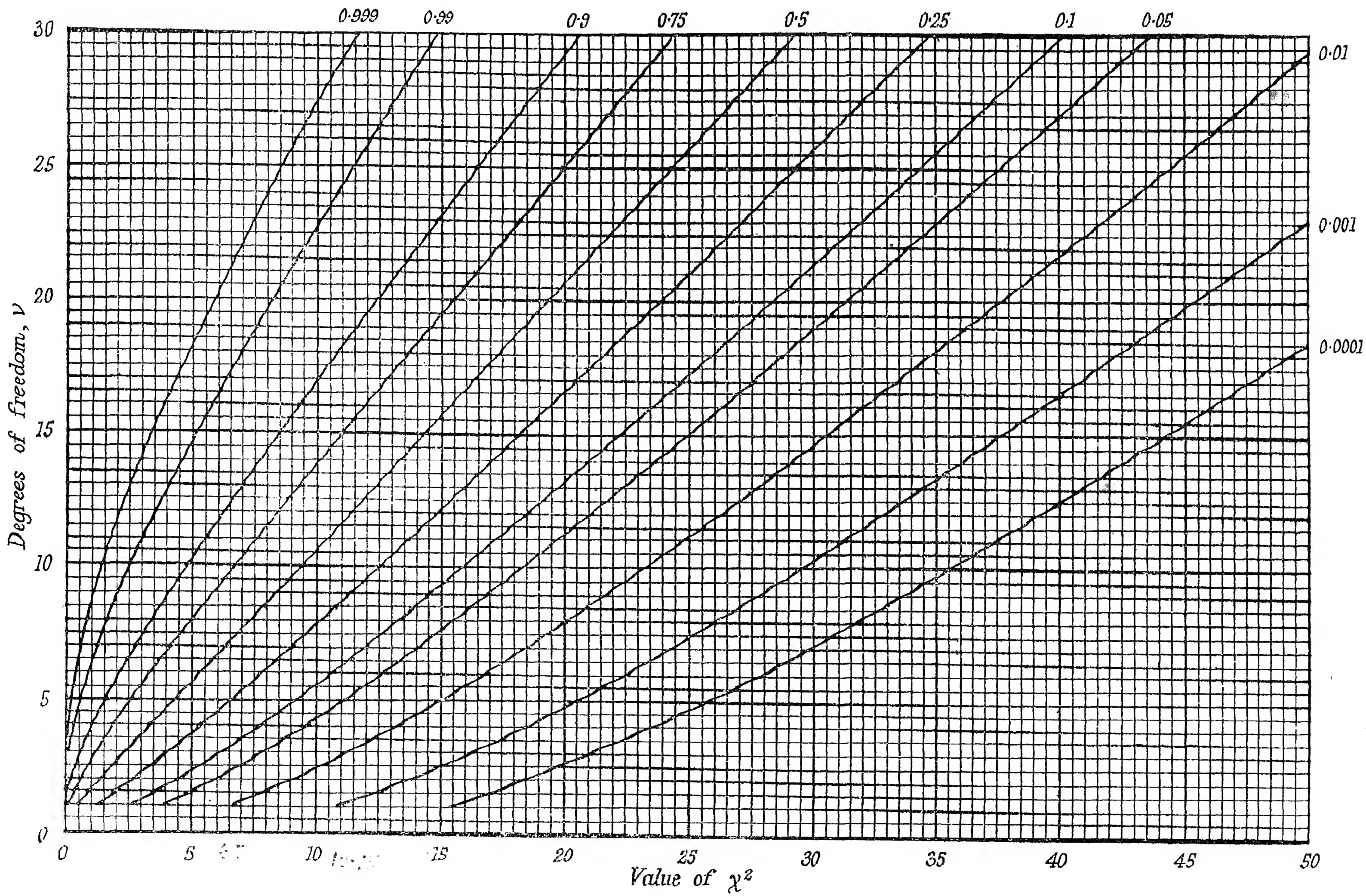
Values of ν_1.

alues of ν_2.	1.	2.	3.	4.	5.			12.	24.	
1	4·1535	4·2585	4·2974	4·3175	4·3297	4·3379	4·3482	4·3585	4·3689	4·3794
2	2·2950	2·2976	2·2984	2·2988	2·2991	2·2992	2·2994	2·2997	2·2999	2·3001
3	1·7649	1·7140	1·6915	1·6786	1·6703	1·6645	1·6569	1·6489	1·6404	1·6314
4	1·5270	1·4452	1·4075	1·3856	1·3711	1·3609	1·3473	1·3327	1·3170	1·3000
5	1·3943	1·2929	1·2449	1·2164	1·1974	1·1838	1·1656	1·1457	1·1239	1·0997
6	1·3103	1·1955	1·1401	1·1068	1·0843	1·0680	1·0460	1·0218	0·9948	0·9643
7	1·2526	1·1281	1·0672	1·0300	1·0048	0·9864	0·9614	0·9335	0·9020	0·8658
8	1·2106	1·0787	1·0135	0·9734	0·9459	0·9259	0·8983	0·8673	0·8319	0·7904
9	1·1786	1·0411	0·9724	0·9299	0·9006	0·8791	0·8494	0·8157	0·7769	0·7305
10	1·1535	1·0114	0·9399	0·8954	0·8646	0·8419	0·8104	0·7744	0·7324	0·6816
11	1·1333	0·9874	0·9136	0·8674	0·8354	0·8116	0·7785	0·7405	0·6958	0·6408
12	1·1166	0·9677	0·8919	0·8443	0·8111	0·7864	0·7520	0·7122	0·6649	0·6061
13	1·1027	0·9511	0·8737	0·8248	0·7907	0·7652	0·7295	0·6882	0·6386	0·5761
14	1·0909	0·9370	0·8581	0·8082	0·7732	0·7471	0·7103	0·6675	0·6159	0·5500
15	1·0807	0·9249	0·8448	0·7939	0·7582	0·7314	0·6937	0·6496	0·5961	0·5269
16	1·0719	0·9144	0·8331	0·7814	0·7450	0·7177	0·6791	0·6339	0·5786	0·5064
17	1·0641	0·9051	0·8229	0·7705	0·7335	0·7057	0·6663	0·6199	0·5630	0·4879
18	1·0572	0·8970	0·8138	0·7607	0·7232	0·6950	0·6549	0·6075	0·5491	0·4712
19	1·0511	0·8897	0·8057	0·7521	0·7140	0·6854	0·6447	0·5964	0·5366	0·4560
20	1·0457	0·8831	0·7985	0·7443	0·7058	0·6768	0·6355	0·5864	0·5253	0·4421
21	1·0408	0·8772	0·7920	0·7372	0·6984	0·6690	0·6272	0·5773	0·5150	0·4294
22	1·0363	0·8719	0·7860	0·7309	0·6916	0·6620	0·6196	0·5691	0·5056	0·4176
23	1·0322	0·8670	0·7806	0·7251	0·6855	0·6555	0·6127	0·5615	0·4969	0·4068
24	1·0285	0·8626	0·7757	0·7197	0·6799	0·6496	0·6064	0·5545	0·4890	0·3967
25	1·0251	0·8585	0·7712	0·7148	0·6747	0·6442	0·6006	0·5481	0·4816	0·3872
26	1·0220	0·8548	0·7670	0·7103	0·6699	0·6392	0·5952	0·5422	0·4748	0·3784
27	1·0191	0·8513	0·7631	0·7062	0·6655	0·6346	0·5902	0·5367	0·4685	0·3701
28	1·0164	0·8481	0·7595	0·7023	0·6614	0·6303	0·5856	0·5316	0·4626	0·3624
29	1·0139	0·8451	0·7562	0·6987	0·6576	0·6263	0·5813	0·5269	0·4570	0·3550
30	1·0116	0·8423	0·7531	0·6954	0·6540	0·6226	0·5773	0·5224	0·4519	0·3481
60	0·9784	0·8025	0·7086	0·6472	0·6028	0·5687	0·5189	0·4574	0·3746	0·2352
	0·9462	0·7636	0·6651	0·5999	0·5522	0·5152	0·4604	0·3908	0·2913	

APPENDIX TABLE 6

Distribution Function of χ^2 for One Degree of Freedom for Values of χ^2 from $\chi^2 = 0$ to $\chi^2 = 1$ by steps of 0·01.

χ^2	P	Δ	χ^2	P	Δ
0	1·00000	7966	0·50	0·47950	436
0·01	0·92034	3280	0·51	0·47514	430
0·02	0·88754	2505	0·52	0·47084	423
0·03	0·86249	2101	0·53	0·46661	418
0·04	0·84148	1842	0·54	0·46243	411
0·05	0·82306	1656	0·55	0·45832	406
0·06	0·80650	1516	0·56	0·45426	400
0·07	0·79134	1404	0·57	0·45026	395
0·08	0·77730	1312	0·58	0·44631	389
0·09	0·76418	1235	0·59	0·44242	384
0·10	0·75183	1169	0·60	0·43858	379
0·11	0·74014	1111	0·61	0·43479	374
0·12	0·72903	1060	0·62	0·43105	369
0·13	0·71843	1015	0·63	0·42736	365
0·14	0·70828	974	0·64	0·42371	360
0·15	0·69854	938	0·65	0·42011	355
0·16	0·68916	905	0·66	0·41656	351
0·17	0·68011	874	0·67	0·41305	346
0·18	0·67137	845	0·68	0·40959	343
0·19	0·66292	820	0·69	0·40616	338
0·20	0·65472	795	0·70	0·40278	334
0·21	0·64677	773	0·71	0·39944	330
0·22	0·63904	752	0·72	0·39614	326
0·23	0·63152	731	0·73	0·39288	322
0·24	0·62421	713	0·74	0·38966	318
0·25	0·61708	696	0·75	0·38648	315
0·26	0·61012	679	0·76	0·38333	311
0·27	0·60333	663	0·77	0·38022	308
0·28	0·59670	648	0·78	0·37714	304
0·29	0·59022	634	0·79	0·37410	301
0·30	0·58388	620	0·80	0·37109	297
0·31	0·57768	607	0·81	0·36812	294
0·32	0·57161	595	0·82	0·36518	291
0·33	0·56566	583	0·83	0·36227	287
0·34	0·55983	572	0·84	0·35940	285
0·35	0·55411	560	0·85	0·35655	281
0·36	0·54851	551	0·86	0·35374	278
0·37	0·54300	540	0·87	0·35096	276
0·38	0·53760	530	0·88	0·34820	272
0·39	0·53230	521	0·89	0·34548	270
0·40	0·52709	512	0·90	0·34278	267
0·41	0·52197	503	0·91	0·34011	264
0·42	0·51694	495	0·92	0·33747	261
0·43	0·51199	487	0·93	0·33486	258
0·44	0·50712	479	0·94	0·33228	256
0·45	0·50233	471	0·95	0·32972	253
0·46	0·49762	463	0·96	0·32719	251
0·47	0·49299	457	0·97	0·32468	248
0·48	0·48842	449	0·98	0·32220	246
0·49	0·48393	443	0·99	0·31974	243
0·50	0·47950	436	1·00	0·31731	241



APPENDIX DIAGRAM.—Contour Lines of the Surface $P = f(\nu, \chi^2)$.

INDEX

(*References are to pages.*)